(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 August 2001 (16.08.2001)

PCT

(10) International Publication Number
WO 01/58923 A2

(51) International Patent Classification[7]: C07K

(21) International Application Number: PCT/US01/00684

(22) International Filing Date: 7 February 2001 (07.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/180,669	7 February 2000 (07.02.2000)	US
09/654,499	1 September 2000 (01.09.2000)	US
09/759,152	16 January 2001 (16.01.2001)	US

(71) Applicant: **TROPIX, INC.** [US/US]; 47 Wiggins Avenue, Bedford, MA 01730 (US).

(72) Inventors: **PALMER, Michelle, A., J.**; 87 Medford Street, Arlington, MA 02174 (US). **GEE, Melissa**; 17 Crescent Avenue, Bedford, MA 01730 (US). **TILLOTSON, Bonnie**; 4 Ripley Road, Belmont, MA 02178 (US). **CHANG, Xiao-jia**; 25 Round Hill Road, Lincoln, MA 01773 (US).

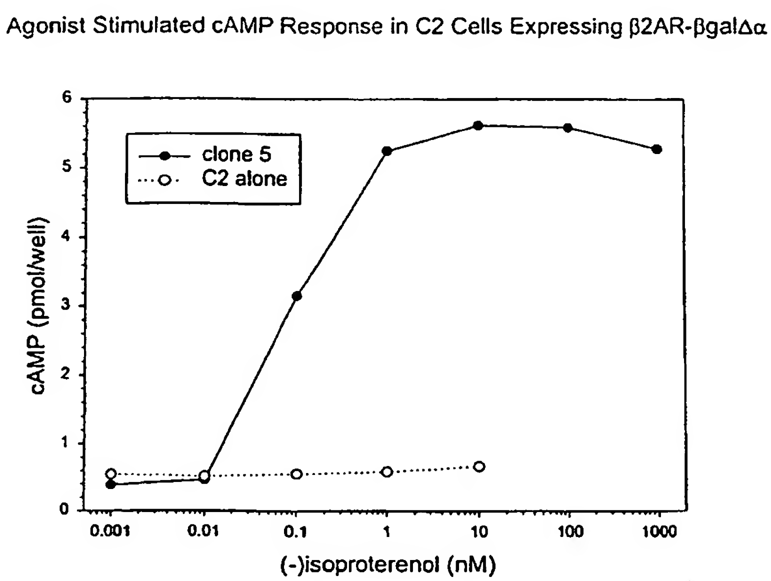
(74) Agents: **KELBER, Steven, B.** et al.; Piper Marbury Rudnick & Wolfe LLP, 1200 Nineteenth Street, N.W., Washington, DC 20036-2412 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: IMPROVED SYSTEMS FOR SENSITIVE DETECTION OF G-PROTEIN COUPLED RECEPTOR AND ORPHAN RECEPTOR FUNCTION USING REPORTER ENZYME MUTANT COMPLEMENTATION



(57) Abstract: Methods for detecting G-protein coupled receptor (GPCR) activity; methods for assaying GPCR activity; and methods for screening for GPCR ligands, G-protein-coupled receptor kinase (GRK) activity, and compounds that interact with components of the GPCR regulatory process are described. Included are methods for expanding ICAST technologies for assaying GPCR activity with applications for ligand fishing, and agonist or antagonist screening. These methods include: engineering seronine/threonine phosphorylation sites into known or orphan GPCR open reading frames in order to increase the affinity of arrestin for the activated form of the GPCR or to increase the reside time of arrestin on the activated GPCR; engineering mutant arrestin proteins that bind to activated GPCRs in the absence of G-protein coupled receptor kinases which may be limiting; and engineering mutant super arrestin proteins that have an increased affinity for activated GPCRs with or without phosphorylation. These methods are intended to increase the robustness of the GPCR/ICAST technology in situations in which G-protein coupled receptor kinases are absent or limiting, or in which the GPCR is not efficiently down-regulated or is rapidly resensitized (thus having a labile interaction with arrestin). Included are also more specific methods for using ICAST complementary enzyme fragments to monitor GPCR homo- and hetero- dimerization with applications for drug lead discovery and ligand and function discovery for orphan GPCRs.

WO 01/58923 A2

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TITLE OF THE INVENTION

IMPROVED SYSTEMS FOR SENSITIVE DETECTION OF G-PROTEIN COUPLED RECEPTOR AND ORPHAN RECEPTOR FUNCTION USING REPORTER ENZYME MUTANT COMPLEMENTATION

BACKGROUND OF THE INVENTION

This application is a continuation-in-part of U.S. Application Serial No. 09/654,499, filed September 1, 2000, which claims the benefit from Provisional Application Serial No. 60/180,669, filed February 7, 2000. The entirety of U.S. Application Serial No. 09/654,499 and Provisional Application Serial No. 60/180,669 are incorporated herein by reference.

Field of the Invention

The present invention relates to methods of detecting G-protein-coupled receptor (GPCR) activity, and provides methods of assaying GPCR activity, methods for screening for GPCR ligands, agonists and/or antagonists, methods for screening natural and surrogate ligands for orphan GPCRs, and methods for screening compounds that interact with components of the GPCR regulatory process.

Background of the Technology

The actions of many extracellular signals are mediated by the interaction of G-protein- coupled receptors (GPCRs) and guanine nucleotide-binding regulatory proteins (G-proteins). G-protein-mediated signaling systems have been identified in many divergent organisms, such as mammals and yeast. The GPCRs represent a

large super family of proteins which have divergent amino acid sequences, but share common structural features, in particular, the presence of seven transmembrane helical domains. GPCRs respond to, among other extracellular signals, neurotransmitters, hormones, odorants and light. Individual GPCR types activate a particular signal transduction pathway; at least ten different signal transduction pathways are known to be activated via GPCRs. For example, the beta 2-adrenergic receptor (β2AR) is a prototype mammalian GPCR. In response to agonist binding, β2AR receptors activate a G-protein (Gs) which in turn stimulates adenylate cyclase activity and results in increased cyclic adenosine monophosphate (cAMP) production in the cell.

The signaling pathway and final cellular response that result from GPCR stimulation depends on the specific class of G-protein with which the particular receptor is coupled (Hamm, "The Many Faces of G-Protein Signaling." J. Biol. Chem., 273:669-672 (1998)). For instance, coupling to the Gs class of G-proteins stimulates cAMP production and activation of the Protein Kinase A and C pathways, whereas coupling to the Gi class of G-proteins down regulates cAMP. Other second messenger systems such as calcium, phospholipase C, and phosphatidylinositol 3 may also be utilized. As a consequence, GPCR signaling events have predominantly been measured via quantification of these second messenger products.

The decrease of a response to a persistent stimulus is a widespread biological phenomenon. Signaling by diverse GPCRs is believed to be terminated by a uniform two-step mechanism. Activated receptor is first phosphorylated by a

GPCR kinase (GRK). An arrestin protein binds to the activated and phosphorylated receptor, thus blocking G-protein interaction. This process is commonly referred to as desensitization, a general mechanism that has been demonstrated in a variety of functionally diverse GPCRs. Arrestin also plays a part in regulating GPCR internalization and resensitization, processes that are heterogenous among different GPCRs (Oakley, et al., J. Biol. Chem., 274:32248-32257 (1999)). The interaction between an arrestin and GPCR in processes of internalization and resensitization is dictated by the specific sequence motif in the carboxyl terminus of a given GPCR. Only a subset of GPCRs, which possess clusters of three serine or threonine residues at the carboxyl termini, were found to co-traffick with the arrestins into the endocytic vesicles after ligand stimulation. The number of receptor kinases and arrestins involved in desensitization of GPCRs is rather limited.

A common feature of GPCR physiology is desensitization and recycling of the receptor through the processes of receptor phosphorylation, endocytosis and dephosphorylation (Ferguson, et al., "G-protein-coupled receptor regulation: role of G-protein-coupled receptor kinases and arrestins." Can. J. Physiol. Pharmacol., 74:1095-1110 (1996)). Ligand-occupied GPCRs can be phosphorylated by two families of serine/threonine kinases, the G-protein-coupled receptor kinases (GRKs) and the second messenger-dependent protein kinases such as protein kinase A and protein kinase C. Phosphorylation by either class of kinases serves to down-regulate the receptor by uncoupling it from its corresponding G-protein. GRK-phosphorylation also serves to down-regulate the receptor by recruitment of a

class of proteins known as the arrestins that bind the cytoplasmic domain of the receptor and promote clustering of the receptor into endocytic vescicles. Once the receptor is endocytosed, it will either be degraded in lysosomes or dephosphorylated and recycled back to the plasma membrane as a fully-functional receptor.

Binding of an arrestin protein to an activated receptor has been documented as a common phenomenon of a variety of GPCRs ranging from rhodopsin to β2AR to the neurotensin receptor (Barak, et al., "A β-arrestin/Green Fluorescent Fusion Protein Biosensor for Detecting G-Protein-Coupled Receptor Activation," J. Biol. Chem., 272:27497-500 (1997)). Consequently, monitoring arrestin interaction with a specific GPCR can be utilized as a generic tool for measuring GPCR activation. Similarly, a single G-protein and GRK also partner with a variety of receptors (Hamm, et al. (1998) and Pitcher et al., "G-Protein-Coupled Receptor Kinases," Annu. Rev. Biochem., 67:653-92 (1998)), such that these protein/protein interactions may also be monitored to determine receptor activity.

Many therapeutic drugs in use today target GPCRs, as they regulate vital physiological responses, including vasodilation, heart rate, bronchodilation, endocrine secretion and gut peristalsis. See, e.g., Lefkowitz et al., Annu. Rev. Biochem., 52:159 (1983). Some of these drugs mimic the ligand for this receptor. Other drugs act to antagonize the receptor in cases when disease arises from spontaneous activity of the receptor.

Efforts such as the Human Genome Project are identifying new GPCRs ("orphan" receptors) whose physiological roles and ligands are unknown. It is estimated that several thousand GPCRs exist in the human genome.

Various approaches have been used to monitor intracellular activity in response to a stimulant, e.g., enzyme-linked immunosorbent assay (ELISA); Fluorescense Imaging Plate Reader assay (FLIPR™, Molecular Devices Corp., Sunnyvale, CA); EVOscreen™, EVOTEC™, Evotec Biosystems Gmbh, Hamburg, Germany; and techniques developed by CELLOMICS™, Cellomics, Inc., Pittsburgh, PA.

Germino et al., "Screening for in vivo protein-protein interactions." Proc. Natl. Acad. Sci., 90(3):933-937 (1993), discloses an *in vivo* approach for the isolation of proteins interacting with a protein of interest.

Phizicky et al., "Protein-protein interactions: methods for detection and analysis." Microbiol. Rev., 59(1): 94-123 (1995), discloses a review of biochemical, molecular biological and genetic methods used to study protein-protein interactions.

Offermanns et al., "$G\alpha_{15}$ and $G\alpha_{16}$ Couple a Wide Variety of Receptors to Phospholipase C." J. Biol. Chem., 270(25):15175-15180 (1995), discloses that $G\alpha_{15}$ and $G\alpha_{16}$ can be activated by a wide variety of G-protein-coupled receptors. The selective coupling of an activated receptor to a distinct pattern of G-proteins is regarded as an important requirement to achieve accurate signal transduction. Id.

Barak et al., "A β-arrestin/Green Fluorescent Protein Biosensor for Detecting G Protein-coupled Receptor Activation." J. Biol. Chem., 272(44):27497-

27500 (1997) and U.S. Patents Nos. 5,891,646 and 6,110,693 disclose the use of a β-arrestin/green fluorescent fusion protein (GFP) for imaging protein translocation upon stimulation of GPCR with optical devices.

Each of the references described above has drawbacks. For example,

- The prior art methodologies require over-expression of the proteins, which could cause artifact and tip the balance of cellular regulatory machineries.
- The prior art visualization or imaging assays are low throughput and lack thorough quantification. Therefore, they are not suitable for high throughput pharmacological and kinetic assays.

In addition, many of the prior art assays require isolation of the GPCR rather than observation of the GPCR in a cell. There thus exists a need for improved methods for monitoring GPCR function.

SUMMARY OF THE INVENTION

The present invention provides modifications to the disclosure in U.S. Application Serial No. 09/654,499. In particular, the present invention is directed to modifications of the below aspects of the invention to further enhance assay sensitivity. The modifications include the use of genetically modified arrestins that exhibit enhanced binding to activated GPCR regardless of whether the GPCR is phosphorylated or non-phosphorylated; the use of a serine/threonine cluster strategy to facilitate screening assays for orphan receptors that do not possess this

structural motif on their own; and the use of a combination of the above modifications to achieve even more enhanced detection.

A first aspect of the present invention is a method that monitors GPCR function proximally at the site of receptor activation, thus providing more information for drug discovery purposes due to fewer competing mechanisms. Activation of the GPCR is measured by a read-out for interaction of the receptor with a regulatory component such as arrestin, G-protein, GRK or other kinases, the binding of which to the receptor is dependent upon agonist occupation of the receptor. The present invention involves the detection of protein/protein interaction by complementation of mutant reporter enzymes.

Binding of arrestin to activated GPCR is a common process in the first step of desensitization that has been demonstrated for most, if not all, GPCRs studied so far. Measurement of GPCR interaction with arrestin via mutant enzyme complementation (i.e., ICAST) provides a more generic assay technology applicable for a wide variety of GPCRs and orphan receptors.

A further aspect of the present invention is a method of assessing GPCR pathway activity under test conditions by providing a test cell that expresses a GPCR, e.g., muscarinic, adrenergic, dopamine, angiotensin or endothelin, as a fusion protein to a mutant reporter enzyme and interacting a protein in the GPCR pathway, e.g., G-protein, arrestin or GRK, as a fusion protein with a complementing mutant reporter enzyme. When test cells are exposed to a known agonist to the target GPCR under test conditions, activation of the GPCR will be

monitored by complementation of the reporter enzyme. Increased reporter enzyme activity reflects interaction of the GPCR with its interacting protein partner.

A further aspect of the present invention is a method of assessing GPCR pathway activity in the presence of a test arrestin, e.g., β-arrestin.

A further aspect of the present invention is a method of assessing GPCR pathway activity in the presence of a test G-protein.

A further aspect of the present invention is a method of assessing GPCR pathway activity upon exposure of the test cell to a test ligand.

A further aspect of the present invention is a method of assessing GPCR activity upon co-expression in the test cell of a second receptor. The second receptor could be the same GPCR or orphan receptor (i.e., homo-dimerization), a different GPCR or orphan receptor (i.e., hetero-dimerization) or could be a receptor of another type.

A further aspect of the present invention is a method for screening for a ligand or agonist to an orphan GPCR. The ligand or agonist could be contained in natural or synthetic libraries or mixtures or could be a physical stimulus. A test cell is provided that expresses the orphan GPCR as a fusion protein with a mutant reporter enzyme, e.g., a β-galactosidase mutant, and, for example, an arrestin or mutant form of arrestin as a fusion protein with a complementing mutant reporter enzyme, e.g., another β-galactosidase mutant. The interaction of the arrestin with the orphan GPCR upon receptor activation is measured by enzymatic activity of the complemented reporter enzyme. The test cell is exposed to a test compound, and an increase in reporter enzyme activity indicates the presence of a ligand or agonist.

A further aspect of the present invention is a method for screening a protein of interest, for example, an arrestin protein (or mutant form of the arrestin protein) for the ability to bind to a phosphorylated, or activated, GPCR. A test cell is provided that expresses a GPCR as a fusion protein with a mutant reporter enzyme, e.g., a β-galactosidase mutant, and contains arrestin (or a mutant form of arrestin) as a fusion protein with a complementing mutant reporter enzyme, e.g., another β-galactosidase mutant. The interaction of arrestin with the GPCR upon receptor activation is measured by enzymatic activity of the complemented reporter enzyme. The test cell is exposed to a known GPCR agonist and then reporter enzyme activity is detected. Increased reporter enzyme activity indicates that the β-arrestin molecule can bind to phosphorylated, or activated, GPCR in the test cell.

A further aspect of the present invention is a method to screen for an agonist to a specific GPCR. The agonist could be contained in natural or synthetic libraries or could be a physical stimulus. A test cell is provided that expresses a GPCR as a fusion protein with a mutant reporter enzyme, e.g., a β-galactosidase mutant, and, for example, an arrestin as a fusion protein with a complementing mutant reporter enzyme, e.g., another β-galactosidase mutant. The interaction of arrestin with the GPCR upon receptor activation is measured by enzymatic activity of the complemented reporter enzyme. The test cell is exposed to a test compound, and an increase in reporter enzyme activity indicates the presence of an agonist. The test cell may express a known GPCR or a variety of known GPCRs, or may express an unknown GPCR or a variety of unknown GPCRs. The GPCR may be, for example, an odorant GPCR or a βAR GPCR.

A further aspect of the present invention is a method for screening a test compound for GPCR antagonist activity. A test cell is provided that expresses a GPCR as a fusion protein with a mutant reporter enzyme, e.g., a β-galactosidase mutant, and, for example, an arrestin as a fusion protein with a complementing mutant reporter enzyme, e.g., another β-galactosidase mutant. The interaction of arrestin with the GPCR upon receptor activation is measured by enzymatic activity of the complemented reporter enzyme. The test cell is exposed to a test compound, and an increase in reporter enzyme activity indicates the presence of an agonist. The cell is exposed to a test compound and to a GPCR agonist, and reporter enzyme activity is detected. When exposure to the agonist occurs at the same time as or subsequent to exposure to the test compound, a decrease in reporter enzyme activity after exposure to the test compound indicates that the test compound has antagonist activity to the GPCR.

A further aspect of the present invention is a method of screening a sample solution for the presence of an agonist, antagonist or ligand to a GPCR. A test cell is provided that expresses GPCR as a fusion protein with a mutant reporter enzyme, e.g., a β-galactosidase mutant, and contains, for example, a β-arrestin as a fusion protein with a complementing reporter, e.g., another β-galactosidase mutant. The test cell is exposed to a sample solution, and reporter enzyme activity is assessed. Changed reporter enzyme activity after exposure to the sample solution indicates the sample solution contains an agonist, antagonist or ligand for a GPCR expressed in the cell.

A further aspect of the present invention is a method of screening a cell for the presence of a GPCR. According to this aspect, an arrestin fusion protein with a mutant reporter enzyme and a GPCR downstream signaling fusion protein with a mutant reporter enzyme are employed to detect GPCR action. A modification of this aspect of the invention can be employed to provide a method of screening a plurality of cells for those cells which contain a GPCR. According to this aspect, a plurality of cells containing a conjugate comprising a β-arrestin protein as a fusion protein with a reporter enzyme are provided; the plurality of cells are exposed to a GPCR agonist; and activity of reporter enzyme activity is detected. An increase in reporter enzymatic activity after exposure to the GPCR agonist indicates β-arrestin protein binding to a GPCR, thereby indicating that the cell contains a GPCR responsive to the GPCR agonist.

A further aspect of the invention is a method for mapping GPCR-mediated signaling pathways. For instance, the system could be utilized to monitor interaction of c-src with β-arrestin-1 upon GPCR activation. Additionally, the system could be used to monitor protein/protein interactions involved in cross-talk between GPCR signaling pathways and other pathways such as that of the receptor tyrosine kinases or Ras/Raf. According to this aspect, a test cell is provided that expresses a GPCR or other related protein with a mutant reporter enzyme, e.g., a β-galactosidase mutant, and contains a protein from another pathway as a fusion protein with a complementing mutant reporter enzyme, e.g., another β-galactosidase mutant. Increased reporter enzymatic activity indicates protein/protein interaction.

A further aspect of the invention is a method for monitoring homo- or hetero- dimerization of GPCRs upon agonist or antagonist stimulation. Increasing evidence indicates that GPCR dimerization is important for biological activity (AbdAlla, et al., "AT1-receptor heterodimers show enhanced G-protein activation and altered receptor sequestration." Nature, 407:94-98 (2000); Bockaert, et al., "Molecular tinkering of G protein-coupled receptors: an evolutionary success." EMBO J. 18:1723-29 (1999)). Jordan, et al., "G-protein-coupled receptor heterodimerization modulates receptor function." Nature, 399:697-700 (1999), demonstrated that two non-functional opioid receptors, κ and δ, heterodimerize to form a functional receptor. Gordon et al., "Dopamine D2 receptor dimers and receptor blocking peptides." Bioch. Biophys. Res. Commun. 227:200-204 (1996), showed different pharmacological properties associated with the monomeric and dimeric forms of Dopamine receptor D2. The D2 receptors exist either as monomers that are selective targets for spiperone or as dimer forms that are targets for nemonapride. Herbert, et al., "A peptide derived from a β2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation." J.B.C. 271:16384-92 (1996), demonstrated that the agonist stimulation was found to stabilize the dimeric state of the receptor, whereas inverse agonists favored the monomeric form. Indeed, the same study showed that a peptide corresponding to the sixth transmembrane domain of the β2-adrenergic receptor inhibited both receptor dimerization and activation. Further, Angers et al., Detection of beta-2-adrenergic receptor dimerization in living cells using bioluminescence resonance energy transfer, Proc. Natl. Acad. Sci. USA, 97(7):3684-3689, discloses the use of

β2-adrenergic receptor fusion proteins (i.e., β2-adrenergic receptor fused to luciferase and β2-adrenergic receptor fused to an enhanced red-shifted green fluorescent protein) to study β2-adrenergic receptor dimerization.

GPCR dimerization in the context of cellular physiology and pharmacology can be monitored in accordance with the invention. For example, β-galactosidase complementation can be measured in test cells that co-express GPCR fusion proteins of β-galactosidase mutant enzymes, e.g., $GPCR_1\Delta\alpha$ and $GPCR_2\Delta\omega$ (FIGURE 27). According to this aspect, the interconversion between monomeric to dimeric forms of the GPCRs or orphan receptors can be measured by mutant reporter enzyme complementation. FIGURE 27 illustrates a test cell co-expressing GPCR or an orphan receptor as a fusion protein with $\Delta\alpha$ form of β-galactosidase mutant (e.g., $GPCR_1\Delta\alpha$), and the same GPCR or orphan receptor as a fusion protein with $\Delta\omega$ form of β-galactosidase mutant (e.g., $GPCR_1\Delta\omega$). Formation of the GPCR homodimer is reflected by formation of an active enzyme, which can be measured by enzyme activity assays, such as the Gal-Screen™ assay. Similarly, hetero-dimerization between two distinct GPCRs, or two distinct orphan receptors, or between one known GPCR and one orphan receptor can be analyzed in test cells co-expressing two fusion proteins, e.g., $GPCR_1\ \Delta\alpha$ and $GPCR_2\Delta\omega$. The increased β-galactosidase activity indicates that the two receptors can form a heterodimer.

A further aspect of the invention is a method of monitoring the interconversion between the monomeric and dimeric form of GPCRs under the influence of agonist or antagonist treatment. The test receptor(s) can be between the same GPCR or orphan receptor (homodimer), or between two distinct GPCRs

or orphan receptors (heterodimer). The increased β-galactosidase activity after treatment with a compound means that the compound binds to and/or stabilizes the dimeric form of the receptor. The decreased β-galactosidase activity after treatment with a compound means that the compound binds to and/or stabilizes the monomeric form of the receptor.

A further aspect of the invention is a method of screening a cell for the presence of a GPCR responsive to a GPCR agonist. A cell is provided that contains protein partners that interact downstream in the GPCR's pathway. The protein partners are expressed as fusion proteins to the mutant, complementing enzyme and are used to monitor activation of the GPCR. The cell is exposed to a GPCR agonist and then enzymatic activity of the reporter enzyme is detected. Increased reporter enzyme activity indicates that the cell contains a GPCR responsive to the agonist.

The present invention involves the use of a combination of proprietary technologies (including ICAST™, Intercistronic Complementation Analysis Screening Technology, Gal-Screen™, etc.) to monitor protein/protein interactions in GPCR signaling. As disclosed in U.S. Application Serial No. 09/654,499, the method of the invention in part involves using ICAST™, which in turn involves the use of two inactive β-galactosidase mutants, each of which is fused with one of two interacting target protein pairs, such as a GPCR and an arrestin. The formation of an active β-galactosidase complex is driven by interaction of the target proteins. In this system, β-galactosidase activity can be detected using, e.g., the Gal-Screen™ assay system, wherein direct cell lysis is combined with rapid

ultrasensitive chemiluminescent detection of β-galactosidase reporter enzyme. This system uses, e.g., a Galacton-*Star*® chemiluminescent substrate for measurement in a luminometer as a read out of GPCR activity.

FIGURE 23 is a schematic depicting the use of the complementation technology in the method of the present invention. FIGURE 23 shows two inactive β-galactosidase mutants that become active when they are forced together by specific interactions between the fusion partners of an arrestin molecule and an activated GPCR or orphan receptor. This assay technology will be especially useful in high throughput screening assays for ligand fishing for orphan receptors, a process called de-orphaning. As illustrated in FIGURE 28, a β-galactosidase fusion protein of an orphan receptor (e.g., $GPCR_{orphan}\Delta\alpha$) is co-expressed in the test cell with a fusion protein of β-arrestin (e.g., β-$Arr\Delta\omega$). When the test cell is subjected to compounds, which could be natural or synthetic, the increased β-galactosidase activity means the compound is either a natural or surrogate ligand for this GPCR. The same assay system can be used to find drug leads for the new GPCRs. The increased β-galactosidase activity in the test cell after treatment indicates the agonist activity of the compound. The decreased β-galactosidase activity in the test cell indicates antagonist activity or inverse agonist activity of the compound. In addition, the method of the invention could be used to monitor GPCR-mediated signaling pathways via other downstream signaling components such as G-proteins, GRKs or the proto-oncogene c-Src.

The invention is achieved in part by using ICAST™ protein/protein interaction screening to map signaling pathways. This technology is applicable to

a variety of known and unknown GPCRs with diverse functions. They include, but are not limited to, the following sub-families of GPCRs:

(a) receptors that bind to amine-like ligands-Acetylcholine muscarinic receptor (M1 to M5), alpha and beta Adrenoceptors, Dopamine receptors (D1, D2, D3 and D4), Histamine receptors (H1 and H2), Octopamine receptor and Serotonin receptors (5HT1, 5HT2, 5HT4, 5HT5, 5HT6, 5HT7);

(b) receptors that bind to a peptide ligand-Angiotensin receptor, Bombesin receptor, Bradykinin receptor, C-C chemokine receptors (CCR1 to CCR8, and CCR10), C-X-C type Chemokine receptors (CXC-R5), Cholecystokinin type A receptor, CCK type receptors, Endothelin receptor, Neurotesin receptor, FMLP-related receptors, Somatostatin receptors (type 1 to type 5) and Opioid receptors (type D, K, M, X);

(c) receptors that bind to hormone proteins-Follic stimulating hormone receptor, Thyrotrophin receptor and Lutropin-choriogonadotropic hormone receptor;

(d) receptors that bind to neurotransmitters-substance P receptor, Substance K receptor and neuropeptide Y receptor;

(e) Olfactory receptors-Olfactory type 1 to type 11, Gustatory and odorant receptors;

(f) Prostanoid receptors-Prostaglandin E2 (EP1 to EP4 subtypes), Prostacyclin and Thromboxane;

(g) receptors that bind to metabotropic substances-Metabotropic glutamate group I to group III receptors;

(h) receptors that respond to physical stimuli, such as light, or to chemical stimuli, such as taste and smell; and

(i) orphan GPCRs-the natural ligand to the receptor is undefined.

Use of the ICAST™ technology in combination with the invention provides many benefits to the GPCR screening process, including the ability to monitor protein interactions in any sub-cellular compartment-membrane, cytosol and nucleus; the ability to achieve a more physiologically relevant model without requiring protein overexpression; and the ability to achieve a functional assay for receptor binding allowing high information content.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1. Cellular expression levels of β2 adrenergic receptor (β2AR) and β-arrestin-2 (βArr2) in C2 clones. Quantification of β-galactosidase (β-gal) fusion protein was performed using antibodies against β-gal and purified β-gal protein in a titration curve by a standardized ELISA assay. Figure 1A shows expression levels of β2AR-βgalΔα clones (in expression vector pICAST ALC). Figure 1B shows expression levels of βArr2-βgalΔω in expression vector pICAST OMC4 for clones 9-3, -7, -9, -10, -19 and –24, or in expression vector pICAST OMN4 for clones 12-4, -9, -16, -18, -22 and -24.

FIGURE 2. Receptor β2AR activation was measured by agonist-stimulated cAMP production. C2 cells expressing pICAST ALC β2AR (clone 5) or parental cells were treated with increasing concentrations of (-)isoproterenol and 0.1mM

IBMX. The quantification of cAMP level was expressed as pmol/well.

FIGURE 3. Interaction of activated receptor β2AR and arrestin can be measured by β-galactosidase complementation. Figure 3A shows a time course of β–galactosidase activity in response to agonist (-)isoproterenol stimulation in C2 expressing β2AR-βgalΔα (β2AR alone, in expression vector pICAST ALC), or a pool of doubly transduced C2 co-expressing β2AR-βgalΔα and βArr2-βgalΔω (in expression vectors pICAST ALC and pICAST OMC and clones isolated from the same pod (43-1, 43-2, 43-7 and 43-8)). Figure 3B shows a time course of β–galactosidase activity in response to agonist (-)isoproterenol stimulation in C2 cells expressing β2AR-βgalΔα alone (in expression vector pICAST ALC) and C2 clones co-expressing β2AR-βgalΔα and βArr1-βgalΔω (in expression vectors ICAST ALC and pICAST OMC).

FIGURE 4. Agonist dose response for interaction of β2AR and arrestin can be measured by β-galactosidase complementation. Figure 4A shows a dose response to agonists (-)isoproterenol and procaterol in C2 cells co-expressing β2AR-βgalΔα and βArr2-βgalΔω fusion constructs. Figure 4B shows a dose response to agonists (-)isoproterenol and procaterol in C2 cells co-expressing β2AR-βgalΔα and βArr1-βgalΔω fusion constructs.

FIGURE 5. Antagonist mediated inhibition of receptor activity can be measured by β-galactosidase complementation in cells co-expressing β2AR-βgalΔα and βArr-βgalΔω. Figure 5A shows specific inhibition with adrenergic

antagonists ICI-118,551 and propranolol of β-galactosidase activity in C2 clones co-expressing β2AR-βgalΔα and βArr2-βgalΔω fusion constructs after incubation with agonist (-)isoproterenol. Figure 5B shows specific inhibition of β-galactosidase activity with adrenergic antagonists ICI-118,551 and propranolol in C2 clones co-expressing β2AR-βgalΔα and βArr1-βgalΔω fusion constructs in the presence of agonist (-)isoproterenol.

FIGURE 6. C2 cells expressing adenosine receptor A2a show cAMP induction in response to agonist (CGS-21680) treatment. C2 parental cells and C2 cells co-expressing A2aR-βgalΔα and βArr1-βgalΔω as a pool or as selected clones (47-2 and 47-13) were measured for agonist-induced cAMP response (pmol/well).

FIGURE 7. Agonist stimulated cAMP response in C2 cells co-expressing Dopamine receptor D1 (D1-βgalΔα) and β-arrestin-2 (βArr2-βgalΔω). The clone expressing βArr2-βgalΔω (Arr2 alone) was used as a negative control in the assay. Cells expressing D1-βgalΔα in addition to βArr2-βgalΔω responded agonist treatment (3-hydroxytyramine hydrochloride at 3 μM). D1(PIC2) or D1(PIC3) designate D1 in expression vector pICAST ALC2 or pICAST ALC4, respectively.

FIGURE 8. Variety of mammalian cell lines can be used to generate stable cells for monitoring GPCR and arrestin interactions. FIGURE 8A, FIGURE 8B and FIGURE 8C show the examples of HEK 293, CHO and CHW cell lines co-expressing adrenergic receptor β2AR and arrestin fusion proteins of β-

galactosidase mutants. The β-galactosidase activity was used to monitor agonist-induced interaction of β2AR and arrestin proteins.

FIGURE 9. Beta-gal complementation can be used to monitor β2 adrenergic receptor homo-dimerization. FIGURE 9A shows β–galactosidase activity in HEK 293 clones co-expressing β2AR-βgalΔα and β2AR-βgalΔω. FIGURE 9B shows a cAMP response to agonist (-)isoproterenol in HEK 293 clones co-expressing β2AR-βgalΔα and β2AR-βgalΔω. HEK293 parental cells were included in the assays as negative controls.

FIGURE 10A. pICAST ALC: Vector for expression of β-galΔα as a C-terminal fusion to the target protein. This construct contains the following features: MCS, multiple cloning site for cloning the target protein in frame with the β-galΔα; GS Linker, (GGGGS)n; NeoR, neomycin resistance gene; IRES, internal ribosome entry site; ColE1ori, origin of replication for growth in E. coli; 5'MoMuLV LTR and 3'MoMuLV LTR, viral promoter and polyadenylation signals from the Moloney Murine leukemia virus.

FIGURE 10B. Nucleotide sequence for pICAST ALC.

FIGURE 11A. pICAST ALN: Vector for expression of β-galΔα as an N-terminal fusion to the target protein. This construct contains the following features: MCS, multiple cloning site for cloning the target protein in frame with the β-galΔα; GS Linker, (GGGGS)n; NeoR, neomycin resistance gene; IRES, internal ribosome entry site; ColE1ori, origin of replication for growth in E. coli;

5'MoMuLV LTR and 3'MoMuLV LTR, viral promoter and polyadenylation signals from the Moloney Murine leukemia virus.

FIGURE 11B. Nucleotide sequence for pICAST ALN.

FIGURE 12A. pICAST OMC: Vector for expression of β-galΔω as a C-terminal fusion to the target protein. This construct contains the following features: MCS, multiple cloning site for cloning the target protein in frame with the β-galΔω ; GS Linker, (GGGGS)n; Hygro, hygromycin resistance gene; IRES, internal ribosome entry site; ColE1ori, origin of replication for growth in E. coli; 5'MoMuLV LTR and 3'MoMuLV LTR, viral promoter and polyadenylation signals from the Moloney Murine leukemia virus.

FIGURE 12B. Nucleotide sequence for pICAST OMC.

FIGURE 13A. pICAST OMN: Vector for expression of β-galΔω as an N-terminal fusion to the target protein. This construct contains the following features: MCS, multiple cloning site for cloning the target protein in frame with the β-galΔω; GS Linker, (GGGGS)n; Hygro, hygromycin resistance gene; IRES, internal ribosome entry site; ColE1ori, origin of replication for growth in E. coli; 5'MoMuLV LTR and 3'MoMuLV LTR, viral promoter and polyadenylation signals from the Moloney Murine leukemia virus.

FIGURE 13B. Nucleotide sequence for pICAST OMN.

FIGURE 14. pICAST ALC βArr2: Vector for expression of β-galΔα as a C-terminal fusion to β-arrestin-2. The coding sequence of human β-arrestin-2 (Genebank Accession Number: NM_004313) was cloned in frame to β-galΔα in a

pICAST ALC vector.

FIGURE 15. pICAST OMC βArr2: Vector for expression of β-galΔω as a C-terminal fusion to β-arrestin-2. The coding sequence of human β-arrestin-2 (Genebank Accession Number: NM_004313) was cloned in frame to β-galΔω in a pICAST OMC vector.

FIGURE 16. pICAST ALC βArr1: Vector for expression of β-galΔα as a C-terminal fusion to β-arrestin-1. The coding sequence of human β-arrestin-1 (Genebank Accession Number: NM_004041) was cloned in frame to β-galΔα in a pICAST ALC vector.

FIGURE 17. pICAST OMC βArr1: Vector for expression of β-galΔω as a C-terminal fusion to β-arrestin-1. The coding sequence of human β-arrestin-1 (Genebank Accession Number: NM_004041) was cloned in frame to β-galΔω in a pICAST OMC vector.

FIGURE 18. pICAST ALC β2AR: Vector for expression of β-galΔα as a C-terminal fusion to β2 Adrenergic Receptor. The coding sequence of human β2 Adrenergic Receptor (Genebank Accession Number: NM_000024) was cloned in frame to β-galΔα in a pICAST ALC vector.

FIGURE 19. pICAST OMC β2AR: Vector for expression of β-galΔω as a C-terminal fusion β2 Adrenergic Receptor. The coding sequence of human β2 Adrenergic Receptor (Genebank Accession Number: NM_000024) was cloned in frame to β-galΔω in a pICAST OMC vector.

FIGURE 20. pICAST ALC A2aR: Vector for expression of β-galΔα as a C-terminal fusion to Adenosine 2a Receptor. The coding sequence of human Adenosine 2a Receptor (Genebank Accession Number: NM_000675) was cloned in frame to β-galΔα in a pICAST ALC vector.

FIGURE 21. pICAST OMC A2aR: Vector for expression of β-galΔω as a C-terminal fusion to Adenosine 2a Receptor. The coding sequence of human Adenosine 2a Receptor (Genebank Accession Number: NM_000675) was cloned in frame to β-galΔω in a pICAST OMC vector.

FIGURE 22. pICAST ALC D1: Vector for expression of β-galΔα as a C-terminal fusion to Dopamine D1 Receptor. The coding sequence of human Dopamine D1 Receptor (Genebank Accession Number: X58987) was cloned in frame to β-galΔα in a pICAST ALC vector.

FIGURE 23. A schematic depicting use of the complementation technology in the method of the invention. FIGURE 23 shows two inactive mutant reporter enzymes that become active when the corresponding fusion partners, GPCR and β-arrestin interact.

FIGURE 24. Vector for expression of a GPCR with inserted seronine/threonine amino acid sequences as a fusion with β-galΔα. The open reading frame of a known or orphan GPCR is engineered to contain additional seronine/threonine sequences, such as SSS (seronine, seronine, seronine), within the C-terminal tail. The engineered GPCR is cloned in frame with β-galΔα in a pICAST ALC vector. The pICAST ALC vector contains the following features:

MCS, multiple cloning site for cloning the target protein in frame with the β-galΔα; GS Linker, (GGGGS)n; NeoR, neomycin resistance gene; IRES, internal ribosome entry site; ColE1ori, origin of replication for growth in E. coli; 5'MoMuLV LTR and 3'MoMuLV LTR, viral promotor and polyadenylation signals from the Moloney Murine leukemia virus.

FIGURE 25. Vector for expression of mutant (R170E) β-arrestin2 as a fusion with β-galΔω. The open reading frame of β-arrestin2 is engineered to contain a point mutation that converts arginine 170 to a glutamate. The mutant β-arrestin2 is cloned in frame with β-galΔω in a pICAST OMC vector. The pICAST OMC vector contains the following features: MCS, multiple cloning site for cloning the target protein in frame with the β-galΔα; GS Linker, (GGGGS)n; Hygro, hygromycin resistance gene; IRES, internal ribosome entry site; ColE1ori, origin of replication for growth in E. coli; 5'MoMuLV LTR and 3'MoMuLV LTR, viral promotor and polyadenylation signals from the Moloney Murine leukemia virus.

FIGURE 26. Phosphorylation insensitive Mutant R170E β-Arrestin2Δω binds to β2ARΔα in Response to Agonist Activation. A parental β2ARΔα C2 cell line was tranduced with the Mutant R170E β-Arrestin2Δω construct. Clonal populations co-expressing the two constructions were plated at 10,000 cells/well in 96 well plates and treated with 10μM (-)isoproterenol, 0.3mM ascorbic acid for the indicated time period. β-galactosidase activity was measured by addition of Tropix Gal-Screen™ assay system substrate (Applied Biosystems) and luminescence was measured using a Tropix TR717™ luminometer (Applied Biosystems). Treatments

were performed in triplicate. For comparison, a clonal cell line (43-8) co-expressing β2ARΔα and wild-type β-Arrestin2Δω was also plated at 10,000 cells/well and given the same agonist treatment regimen. Minutes of (-)isoproterenol treatment is shown on the X-axis and β-galactosidase activity indicated by relative light units (RLU) is shown on the Y-axis.

FIGURE 27. GPCR dimerization measured by β-galactosidase complementation. A schematic depicting the utilization of the invention for monitoring GPCR homo- or hetero- dimerization. One GPCR is fused to one complement enzyme fragment, while the second GPCR is fused to the second complement enzyme fragment. Interaction of the two GPCRs is monitored by complementation of the enzyme fragments to produce an active enzyme complex (i.e., β-galactosidase activity). GPCR homo- or hetero- dimerization can be monitored in the absence or presence of ligand, agonists, inverse agonists or antagonists.

FIGURE 28. Ligand fishing for orphan receptors by β-galactosidase mutant complementation in ICAST™ system. A schematic depicting the utilization of the invention for ligand fishing and agonist/antagonist screening for orphan GPCRs. As an example, a test cell expressing two β-gal fusion proteins, $GPCR_{orphan}$-Δα and Arrestin-Δω, is subjected to treatments with samples from natural or synthetic compound libraries, or from tissue extracts, or from conditioned media of cultured cells. An increased β-gal activity after treatment indicates the activation of the orphan receptor by a ligand in the testing sample. The readout of increased β-gal activity reflects the interaction of an activated

GPCR orphan receptor with a β-arrestin. Therefore, a cognate or a surrogate ligand for the testing receptor is identified.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

The present invention provides a method to interrogate GPCR function and pathways. The G-protein-coupled superfamily continues to expand rapidly as new receptors are discovered through automated sequencing of cDNA libraries or genomic DNA. It is estimated that several thousand GPCRs may exist in the human genome. Only a portion have been cloned and even fewer have been associated with ligands. The means by which these, or newly discovered orphan receptors, will be associated with their cognate ligands and physiological functions represents a major challenge to biological and biomedical research. The identification of an orphan receptor generally requires an individualized assay and a guess as to its function. The present invention involves the interrogation of GPCR function by monitoring the activation of the receptor using activation dependent protein-protein interactions between the test GPCR or orphan receptor and a β-arrestin. The specific protein-protein interactions are measured using the mutant enzyme complementation technology disclosed herein. This assay system eliminates the prerequisite guessing because it can be performed with and without prior knowledge of other signaling events. It is sensitive, rapid and easily performed and is applicable to nearly all GPCRs because the majority of these receptors desensitize by a common mechanism.

The present invention provides a complete assay system for monitoring

protein-protein interactions in GPCR pathways. The invention employs the complementation technology, ICAST™ (Intercistronic Complementation Analysis Screening Technology as disclosed in pending U.S. patent application serial no. 053,614, filed April 1, 1998, the entire contents of which are incorporated herein by reference). The ICAST™ technology involves the use of two mutant forms of a reporter enzyme fused to proteins of interest. When the proteins of interest do not interact, the reporter enzyme remains inactive. When the proteins of interest do interact, the reporter enzyme mutants come together and form an active enzyme. According to an embodiment of the invention, the activity of β-galactosidase may be detected with the Gal-Screen™ assay system developed by Advanced Discovery Sciences™, which involves the use of Galacton-*Star*®, an ultrasensitive chemiluminescent substrate. The Gal-Screen™ assay system and the Galacton-*Star*® chemiluminescent substrate are disclosed in U.S. Patent Nos. 5,851,771; 5,538,847; 5,326,882; 5,145,772; 4,978,614; and 4,931,569, the contents of which are incorporated herein by reference in their entirety. The invention provides an array of assays, including GPCR binding assays, that can be achieved directly within the cellular environment in a rapid, non-radioactive assay format. The methods of the invention are an advancement over the invention disclosed in U.S. Patent Nos. 5,891,646 and 6,110,693 and the method disclosed in Angers et al., supra., which rely on microscopic imaging or spectrometry of GPCR components as fusion with Green-fluorescent-protein. The imaging technique disclosed in U.S. Patent Nos. 5,891,646 and 6,110,693 and spectrometry-based technique in Angers et al. are limited by low-throughput and lack of thorough quantification.

The assay system of the invention combined with Advanced Discovery Sciences™ technologies provide highly sensitive cell-based methods for interrogating GPCR pathways which are amenable to high-throughput screening (HTS). Among some of the technologies developed by Advanced Discovery Sciences™ that may be used with the present invention are the Gal-Screen™ assay system (discussed above) and the cAMP-Screen™ immunoassay system. The cAMP-Screen™ immunoassay system provides ultrasensitive determination of cAMP levels in cell lysates. The cAMP-Screen™ assay utilizes the high-sensitivity chemiluminescent alkaline phosphatase (AP) substrate CSPD® (disodium 3-(4-methoxyspiro {1,2-dioxetane-3,2'-(5'-chloro) tricyclo 3.3.1.1.^{3,7}}decan-4-yl phenyl phosphate) with Sapphire-II™ luminescence enhancer.

Unlike yeast-based-two-hybrid assays used to monitor protein/protein interactions in high-throughput assays, the present invention (1) is applicable to a variety of cells including mammalian cells, plant cells, protozoa cells such as E. coli and cells of invertebrate origin such as yeast, slime mold (*Dictyostelium*) and insects; (2) detects interactions at the membrane at the site of the receptor target or in the cytosol at the site of downstream target proteins rather than a limited cellular localization, i.e., nucleus; and (3) does not rely on indirect read-outs such as transcriptional activation. The present invention thus provides assays with greater physiological relevance and fewer false positives.

The present inventors have developed modifications to the embodiment disclosed in U.S. patent application serial no. 053,614 described above in order to enhance the sensitivity of the inventive GPCR assay. According to an

embodiment, the invention incorporates the use of serine/threonine clusters to enhance and prolong the interaction of GPCR with arrestin in order to make the detection more robust. The clusters can be utilized for orphan receptors or known GPCRs, which do not have this sequence motif. By adding this sequence to the C-terminal tail of the receptor, the activation of the receptor can be detected more readily by readouts of arrestin binding to GPCR, i.e., β-galactosidase complementation from fusion proteins of target proteins with β-galactosidase mutants.

According to another embodiment, the invention incorporates the use of arrestin point mutations to bypass the requirement of phosphorylation, by the action of specific GRK, on the C-terminal tail or intracellular loops of GPCR upon activation. The applications include i) wherein the cognate GRK for a particular GPCR or orphan receptor is unknown; and ii) wherein the specific GRK for the receptor of interest (or under test) may not be present or may have low activity in the host cell that is used for receptor activation assay.

According to another embodiment, the invention incorporates the use of a super arrestin to increase the binding efficiency of arrestin to an activated GPCR and to stabilize the GPCR/arrestin complex during GPCR desensitization. This application can be used to increase the robustness of ICAST/GPCR applications in cases where the GPCR is normally resensitized rapidly post desensitization.

Each of these methodologies is discussed below.

The invention will now be described in the following non-limiting examples.

EXAMPLE:

According to an embodiment of the invention, GPCR activation is measured through monitoring the binding of arrestin to ligand-activated GPCR. In this assay system, a GPCR, e.g., β-adrenergic receptor (β2AR), and an arrestin, e.g., β-arrestin, are co-expressed in the same cell as fusion proteins with mutant forms of a reporter enzyme, e.g., β-galactosidase (β-gal). As illustrated in Figure 23, the β2AR is expressed as a fusion protein with Δα form of β-gal mutant (β2ARΔα) and the β-arrestin as a fusion protein with the Δω form of β-gal mutant (β-ArrΔω). The two fusion proteins, which at first exist in a resting (or un-stimulated) cell in separate compartments, i.e., the membrane for GPCR and the cytosol for arrestin, cannot form an active β-galactosidase enzyme. When such a cell is treated with an agonist or a ligand, the ligand-occupied and activated receptor becomes a high affinity binding site for arrestin. The interaction between an activated GPCR, β2ARΔα, and arrestin, β-ArrΔω, drives the β-gal mutant complementation. The enzyme activity can be measured by using an enzyme substrate, which upon cleavage releases a product measurable by colorimetry, fluorescence, or chemiluminescence (e.g., the Gal-Screen™ assay system).

Experiment protocol-

1. In the first step, the expression vectors for β2ARΔα and βArr2Δω were engineered in selectable retroviral vectors pICAST ALC, as described in Figure 18 and pICAST OMC, as described in Figure 15.

2. In the second step, the two expression constructs were transduced into either C2C12 myoblast cells, or other mammalian cell lines, such as COS-7, CHO, A431, HEK 293, and CHW. Following selection with antibiotic drugs, stable clones expressing both fusion proteins at appropriate levels were selected.

3. In the last step, the cells expressing both β2ARΔα and βArr2Δω were tested for response by agonist/ligand stimulated β-galactosidase activity. Triplicate samples of cells were plated at 10,000 cells in 100 microliter volume into a well of 96-well culture plate. Cells were cultured for 24 hours before assay. For agonist assay (Figures 3 and 4), cells were treated with variable concentrations of agonist, for example, (-) isoproterenol, procaterol, dobutamine, terbutaline or L-L-phenylephrine for 60 min at 37° C. The induced β- galactosidase activity was measured by addition of Tropix Gal-Screen™ assay system substrate (Applied Biosystems) and luminescence measured in a Tropix TR717™ luminometer (Applied Biosystems). For antagonist assay (Figure 5), cells were pre-incubated for 10 min in fresh medium without serum in the presence of ICI-118,551 or propranolol followed by addition of 10 micro molar (-) isoproterenol.

Serine/Threonine Cluster Strategy

Background

Based on structure-function relationship studies on β-arrestins, a large region within the amino-terminal half of β-arrestins (termed the activation-recognition domain) recognizes the agonist-activated state of GPCRs. This region of β-arrestin also contains a small positively charged domain (approximately 20

amino acids with net charge +7) called the phosphorylation-recognition domain, which appears to interact with the GRK-phosphorylated carboxyl termini of GPCRs.

GPCRs can be divided into two classes based on their affinities for β-arrestins. Oakley et al., "Association of β-Arrestin with G Protein-Coupled Receptors During Clathrin-Mediated Endocytosis Dictates the Profile of Receptor Resensitization." J. Biol. Chem., 274(45):32248-32257 (1999). The molecular determinants underlying this classification appear to reside in specific serine or threonine residues located in the carboxyl-terminal tail of the receptor. The receptor class that contains serine/threonine clusters (defined as serine or threonine residues occupying three consecutive or three out of four positions) in the carboxyl-termini binds β-arrestin with high affinity upon activation and phosphorylation and remains bound with β-arrestin even after receptor internalization, whereas the receptor class that contains only scattered serine and threonine residues in the carboxy-terminal tail binds β-arrestins with less affinity and disassociates from the β-arrestin upon internalization. Several known GPCRs, such as vasopressin V2 receptor (Oakley, et al.), neurotensin receptor 1 and angiotensin II receptor type 1A (Zhang, et al., "Cellular Trafficking of G Protein-Coupled Receptor/β-Arrestin Endocytic Complexes." J. Biol. Chem., 274(16):10999-11006 (1999)), which possess one or more of such serine/threonine clusters in their carboxyl-termini, were shown to bind β-arrestins with high affinity.

EXAMPLE

According to an embodiment of the invention, a serine/threonine cluster strategy is used to facilitate screening assays for orphan receptors that do not possess this structural motif of their own. The orphan receptors are easily classified by sequence alignment. Orphan receptors lacking the serine/threonine clusters are each cloned into an expression vector that is modified to introduce one or more serine/threonine cluster(s) to the carboxyl-terminal tail of the receptor (FIGURE 24). The serine/threonine clusters enhance the receptor activation dependent interaction between the activated and phosphorylated receptor (negative charges) and β-arrestin (positive charges in the phosphorylation-recognition domain) through strong ionic interactions, thus prolonging interaction between the receptor and arrestin. The modification of the orphan receptor tail thus makes detection of receptor activation more robust.

Experiment protocol -

1. In a first step, the open-reading-frame (ORF) of an orphan receptor, which lacks the serine/threonine clusters, is cloned into a modified expression vector such as pICAST ALC described in Figure 10A. The modified pICAST ALC includes coding sequences for one or more sets of serine/threonine clusters (for example, SSS or SST) located downstream from the insert of the ORF of an orphan receptor (FIGURE 24).

2. In a second step, chimeric orphan receptor, $ORF_{orphan\ R}$-$(SSS)_n$-$\Delta\alpha$, is co-

expressed in a mammalian cell with a β-arrestin chimera, such as βArr2Δω described in Figure 15.

3. In a third step, the cell is treated with an agonist or a ligand and the activated receptor with phosphorylated serine cluster(s) binds the β-arrestin with high affinity producing strong signals in readouts of β-gal complementation.

This assay, which provides a means for sensitive measurement of functional activation of the orphan receptors, can be used to screen for natural or surrogate ligands for orphan receptors, a process called de-orphaning or target discovery for new GPCRs (FIGURE 28). Furthermore, this assay is also useful in screening for potential agonists and antagonists for lead discovery of GPCRs.

<u>Enhanced Binding of Arrestin in the Presence and in the Absence of GPCR Phosphorylation</u>

<u>Background</u>

Six different classes of G-protein coupled receptor kinases (GRKs) have been identified and each of these has been reported to be expressed as multiple splice variants. <u>Krupnick et al.</u>, "The role of receptor kinases and arrestins in G protein-coupled receptor regulation." Ann. Rev. Pharmacol. Toxicol., 38:289-319 (1998). Although many cell lines express a variety of GRKs, the specific GRK required for phosphorylation of a given GPCR may not always be present in the cell line used for recombinant GPCR and arrestin expression. This is particularly an issue for applications using orphan receptors, in which case the cognate GRK will likely be unknown. In other cases, the cell line used for recombinant

expression work may have the required GRK, but may express the GRK at low levels. In order to bypass such caveats, genetically modified arrestins that bind specifically to activated GPCRs, but without the requirement of GRK phosphorylation are employed.

Mutagenesis studies on arrestins demonstrate that point mutations in the phosphorylation-recognition domain, particularly mutations converting Arg175 (of visual arrestin) to an oppositely charged residue such as glutamate (R175E mutation), result in an arrestin which specifically binds to activated GPCRs, but does so without the requirement for phosphorylation.

Numerous observations have led to the hypothesis that arrestin exists in an inactive state that has a low affinity for GPCRs. Once a GPCR is both activated and phosphorylated, the phosphorylated region of the GPCR C-terminus interacts with the phosphorylation-recognition domain of arrestin causing the arrestin to change conformations allowing the activation-recognition region to be exposed for binding to the activated/ phosphorylated receptor. Vishnivetskiy et al., "How does arrestin respond to the phosphorylated state of rhodopsin?" J. Biol. Chem., 274(17):11451-11454 (1999); Gurevich et al., "Arrestin interactions with G protein-coupled receptors. Direct binding studies of wild-type and mutant arrestins with rhodopsin, beta 2-adrenergic and m2 muscarinic cholinergic receptors." J. Biol. Chem., 270(2):720-731, (1995); Gurevich et al., "Mechanism of phosphorylation-recognition by visual arrestin and the transition of arrestin into a high affinity binding site." Mol. Pharmacol., 51(1):161-169 (1997); Kovoor et al., "Targeted construction of phosphorylation-independent beta-arrestin mutants with

constitutive activity in cells." J. Biol. Chem., 274(11):6831-6834 (1999). In summary, binding studies of single mutation, double mutation, deletion, and chimerical arrestins with inactive, inactive and phosphorylated, activated but not phosphorylated, or activated and phosphorylated visual or non-visual GPCRs all support this model.

EXAMPLE

A phosphorylation insensitive mutant of arrestin fused to mutant reporter protein can be produced that will bind to activated GPCRs in a phosphorylation independent manner. As proof of concept, a point mutation for β-arrestin2, R170E β-arrestin2, has been produced and its interaction with β2AR has been analyzed in accordance with the invention.

Experimental protocol:

1) In the first step, β-arrestin2 was mutated such that Arg170 was converted to Glu. This mutation is equivalent to the R175E mutation of visual arrestin. The mutant β-arrestin2 open reading frame was cloned in frame with Δω-β-galactosidase in the pICAST OMC expression vector to produce a modified expression vector R170E β-arrestin2 (FIGURE 25).

2) In the second step, the R170E β-arrestin2 expression construct was transduced into a C2C12 myoblast cell line that had been engineered to express β2AR as a fusion to Δα-β-galactosidase as described in Figure 18 of U.S. Application Serial No. 09/654,499. Following selection with antibiotic drugs, a

population of clones expressing both fusion proteins was obtained.

3) In the last step, this population of cells expressing both R170E β-arrestin2Δω and β2ARΔα were tested for response by agonist/ligand stimulated β-galactosidase activity as demonstrated in FIGURE 26. The C2C12 clone 43-8 co-expressing β2ARΔα and wild-type β-arrestin2Δω (FIGURE 26) was used as reference control. Triplicate samples of cells were plated at 10,000 cells in 100 microliter volume into wells of a 96-well culture plate. Cells were cultured for 24 hours before assay. For agonist assay as in FIGURE 26, cells were treated with 10μm (-)isoproterenol stabilized with 0.3mM ascorbic acid 37° C for 0, 5, 10, 15, 30, 45 or 60 minutes. The induced β-galactosidase activity was measured by addition of Tropix Gal-Screen™ assay system substrate (Applied Biosystems) and luminescence measured in a Tropix TR717™ luminometer (Applied Biosystems). As shown in Figure 26, the mutant arrestin interacts with β2AR in an agonist-dependent manner and was comparable with that of wild-type arrestin.

4) To expand the application of phosphorylation-insensitive arrestin, cell lines such as C2C12, CHO or HEK 293, are developed that express the R170E β-arrestin2Δω construction. These cell lines can be used to transduce orphan or known GPCRs as fusions with Δα-β-galactosidase in order to develop cell lines for agonist and antagonist screening and

Development of Super Arrestins:

Background

Attenuation of GPCR signaling by the arrestin pathway serves to ensure that a cell or organism does not over-react to a stimulus. At the same time, the arrestin pathway often serves to recycle the GPCR such that it can be temporarily inactivated but then quickly resensitized to allow for sensitivity to new stimuli. The down-regulation process involves phosphorylation of the receptor, binding to arrestin and endocytosis. Following endocytosis of the desensitized receptor, the receptor is either degraded in lysosomes or resensitized and sent back to the membrane. Resensitization involves release of arrestin from the receptor, dephosphorylation and cycling back to the membrane. The actual route a GPCR follows upon activation depends on its biological function and the needs of the organism. Because of these diverse pathways that may be required of the down-regulation pathway, arrestin affinities for activated GPCRs vary from receptor to receptor. It would thus be very advantageous to engineer super arrestins that have a higher affinity and avidity for activated GPCRs than what nature has provided.

Although mutational, deletion and chimerical studies of arrestins have focused on understanding regulatory switches in the molecule that respond to GPCR phosphorylation states, several of these altered recombinant forms of arrestin have resulted in molecules with enhanced binding to activated, phosphorylated GPCRs. Conversion of Arg175 to histidine, tyrosine, phenylalanine or threonine results in significantly higher amounts of binding to phosphorylated, activated rhodopsin than wild-type arrestin or R175E arrestin,

although these mutations result in less binding to activated, non-phosphorylated receptor. Gurevich et al. (1997). In addition, conversion of Valine 170 to alanine increased the constitutive affect of the R175E mutation, but also nearly doubled the amount of interaction of wild-type arrestin with activated, phosphorylated rhodopsin. Gurevich et al. (1997).

Truncation of β-arrestin1 at amino acid 382 has been reported to enhance binding of both R169E (equivalent to arrestin R175E) and wild-type β-arrestin1 to activated or activated and phosphorylated receptor, respectively. Kovoor et al. Chimerical arrestins in which functional regions of visual arrestin were swapped with those of β-arrestin1 have been reported to be altered in binding affinity to activated, phosphorylated GPCRs. Gurevich et al. (1995). Several of these chimeras, such as β-arrestin1 containing the visual arrestin extreme N-terminus, show increased specific binding to phosphorylated activated GPCRs compared to wild-type β-arrestin1 (Gurevich et al. (1995)). Modifications that enhance arrestin affinity for the activated GPCR such as described above, whether phosphorylated or non-phosphorylated, could also enhance signal to noise of β-galactosidase activity since the arrestin/GPCR complex is stabilized and/or more long-lived. The use of mutant arrestins with higher activated-GPCR affinity would improve the inventive technology for GPCR targets, without compromising receptor/ligand biology.

In addition, this "super arrestin" approach can be combined with the use of arrestin point mutations to provide a stronger signal to noise with or without GRK requirements.

EXAMPLE

An arrestin mutant fused to mutant reporter protein can be produced to enhance binding of the arrestin to an activated GPCR to enhance sensitivity of detection.

Experiment protocol -

1) In the first step, mutant β-arrestin2 constructions will be generated which include R170E/T/Y/or H, V165A, substitution of a.a. 1-43with a.a. 1-47 of visual arrestin, or deletion of the C-terminal and combinations of these alterations. The mutant β-arrestin2 open reading frames will be cloned in frame with Δω-β-galactosidase in the pICAST OMC expression vector similar to cloning of the R170E β-arrestin2 mutation shown in FIGURE 25.

2) In the second step, mutant expression constructs will be transduced into a C2C12 myoblast cell line that has been engineered to express β2AR as a fusion to Δα-β-galactosidase. Following selection with antibiotic drugs, a population of clones expressing both fusion proteins will be obtained. Wild type and R170E β-arrestin2 constructions will be transduced to generate control, reference clonal populations.

3) In the third step, populations of cells expressing both β-arrestin2Δω (mutant or wild type) and β2ARΔα will be tested for response by agonist/ligand stimulated β-galactosidase activity.

4) In the next step, mutant (super) β-arrestin2Δω constructions that show a significantly higher signal to noise ratio in the agonist assay compared with wild-type β-arrestin2Δω will be chosen. These constructions will be used to develop

stable cell lines expressing the "super" β-arrestin2Δω that can be used for transducing in known or orphan GPCRs. Use of a super β-arrestin2Δω could increase the signal to noise of ICAST/GPCR applications allowing improved screening capabilities for lead and ligand discovery.

Super Arrestin is used to increase the binding efficiency of arrestin to an activated GPCR and to stabilize the GPCR/arrestin complex during GPCR desensitization. This application can be used to increase the robustness of ICAST/GPCR applications in cases where the GPCR is normally resensitized rapidly post desensitization.

The assays of this invention, and their application and preparation have been described both generically, and by specific example. The examples are not intended as limiting. Other substituent identities, characteristics and assays will occur to those of ordinary skill in the art, without the exercise of inventive faculty. Such modifications remain within the scope of the invention, unless excluded by the express recitation of the claims advanced below.

WHAT IS CLAIMED IS:

1. A method of assessing the effect of a test condition on G-protein-coupled receptor (GPCR) pathway activity, comprising:

a) providing a cell that expresses a GPCR as a fusion protein to one mutant form of reporter enzyme and an interacting protein partner as a fusion to another mutant form of enzyme,

wherein said cell also expresses an arrestin, wherein said arrestin is modified to enhance binding of said arrestin to said GPCR, wherein said enhanced binding between said arrestin and said GPCR increases sensitivity of detection of said effect of said test condition;

b) exposing the cell to a ligand for said GPCR under said test condition; and

c) monitoring activation of said GPCR by complementation of said reporter enzyme;

wherein increased reporter enzyme activity in the cell compared to that which occurs in the absence of said test condition indicates increased GPCR interaction with its interacting protein partner compared to that which occurs in the absence of said test condition, and decreased reporter enzyme activity in the cell compared to that which occurs in the absence of said test condition indicates decreased GPCR interaction with its interacting protein partner compared to that which occurs in the absence of said test condition.

2. A method of assessing the effect of a test condition on G-protein-coupled receptor (GPCR) pathway activity, comprising:

a) providing a cell that expresses a GPCR as a fusion protein to one mutant

form of reporter enzyme and an interacting protein partner as a fusion to another mutant form of enzyme;

wherein said GPCR fusion protein is modified to include one or more sets of serine/threonine clusters, wherein said one or more sets of serine/threonine clusters enhance binding of said GPCR to arrestin, wherein said enhanced binding between said GPCR and said arrestin increases sensitivity of detection of said effect of said test condition;

b) exposing the cell to a ligand for said GPCR under said test condition; and

c) monitoring activation of said GPCR by complementation of said reporter enzyme;

wherein increased reporter enzyme activity in the cell compared to that which occurs in the absence of said test condition indicates increased GPCR interaction with said interacting protein partner compared to that which occurs in the absence of said test condition, and decreased reporter enzyme activity in the cell compared to that which occurs in the absence of said test condition indicates decreased GPCR interaction with interacting protein partner compared to that which occurs in the absence of said test condition.

3. A DNA molecule comprising a sequence encoding a biologically active hybrid GPCR, wherein said hybrid GPCR comprises a GPCR as a fusion protein to one mutant form of reporter enzyme and wherein said hybrid GPCR is modified to include one or more sets of serine/threonine clusters, wherein said one or more sets of serine/threonine clusters enhance binding of said hybrid GPCR to arrestin.

4. A DNA construct capable of directing the expression of a biologically

active hybrid GPCR in a cell, comprising the following operatively linked elements:

a promoter; and

a DNA molecule comprising a sequence encoding a biologically active hybrid GPCR, wherein said hybrid GPCR comprises a GPCR as a fusion protein to one mutant form of reporter enzyme and wherein said hybrid GPCR is modified to include one or more sets of serine/threonine clusters, wherein said one or more sets of serine/threonine clusters enhance binding of said hybrid GPCR to arrestin.

5. A cell transformed with a DNA construct capable of expressing a biologically active hybrid GPCR in a cell, comprising the following operatively linked elements:

a promoter; and

a DNA molecule comprising a sequence encoding a biologically active hybrid GPCR, wherein said hybrid GPCR comprises a GPCR as a fusion protein to one mutant form of reporter enzyme and wherein said hybrid GPCR is modified to include one or more sets of serine/threonine clusters, wherein said one or more sets of serine/threonine clusters enhance binding of said hybrid GPCR to arrestin.

6. A DNA molecule comprising a sequence encoding a biologically active hybrid arrestin, wherein said hybrid arrestin comprises an arrestin as a fusion to one mutant form of reporter enzyme and wherein said hybrid arrestin is modified to enhance binding of said arrestin to GPCR.

7. A DNA construct capable of directing the expression of a biologically active hybrid arrestin in a cell, comprising the following operatively linked

elements:

a promoter; and

a DNA molecule comprising a sequence encoding a biologically active hybrid arrestin, wherein said hybrid arrestin comprises an arrestin as a fusion to one mutant form of reporter enzyme and wherein said hybrid arrestin is modified to enhance binding of said arrestin to GPCR.

8. A cell transformed with a DNA construct capable of expressing a biologically active hybrid arrestin in a cell, comprising the following operatively linked elements:

a promoter; and

a DNA molecule comprising a sequence encoding a biologically active hybrid arrestin, wherein said hybrid arrestin comprises an arrestin as a fusion to one mutant form of reporter enzyme and wherein said hybrid arrestin is modified to enhance binding of said arrestin to GPCR.

9. A method of assessing the effect of a test condition on G-protein-coupled receptor (GPCR) pathway activity, comprising:

a) providing a cell that expresses a GPCR as a fusion protein to one mutant form of reporter enzyme and an interacting protein partner as a fusion to another mutant form of enzyme,

wherein said cell also expresses an arrestin, wherein said arrestin is modified by introducing a point mutation in a phosphorylation-recognition domain to remove a requirement for phosphorylation of said GPCR for arrestin binding to permit binding of said arrestin to said GPCR in said cell regardless of whether said

GPCR is phosphorylated,

b) exposing the cell to a ligand for said GPCR under said test condition; and

c) monitoring activation of said GPCR by complementation of said reporter enzyme;

wherein increased reporter enzyme activity in the cell compared to that which occurs in the absence of said test condition indicates increased GPCR interaction with its interacting protein partner compared to that which occurs in the absence of said test condition, and decreased reporter enzyme activity in the cell compared to that which occurs in the absence of said test condition indicates decreased GPCR interaction with its interacting protein partner compared to that which occurs in the absence of said test condition.

10. The method of Claim 9, wherein said arrestin is mutated to increase a property selected from affinity and avidity for activated, non-phosphorylated GPCR.

11. The method of Claim 10, wherein said arrestin is β-arrestin2 and wherein said β-arrestin2 is mutated to convert Arg169 to an oppositely charged residue.

12. The method of Claim 11, wherein said oppositely charged residue is selected from the group consisting of histidine, tyrosine, phenylalanine and threonine.

13. The method of Claim 9, wherein said arrestin is mutated to increase a property selected from affinity and avidity for activated and phosphorylated GPCR.

14. A method of assessing the effect of a test condition on G-protein-

coupled receptor (GPCR) pathway activity, comprising:

a) providing a cell that expresses a GPCR as a fusion protein to one mutant form of reporter enzyme and an interacting protein partner as a fusion to another mutant form of enzyme;

wherein said GPCR fusion protein is modified to include one or more sets of serine/threonine clusters, said one or more serine/threonine clusters defined as serine or threonine residues occupying three consecutive or three out of four positions in a carboxyl-termini of said GPCR, wherein said one or more sets of serine/threonine clusters enhance binding of said GPCR to arrestin, wherein said enhanced binding between said GPCR and said arrestin increases sensitivity of detection of said effect of said test condition;

b) exposing the cell to a ligand for said GPCR under said test condition; and

c) monitoring activation of said GPCR by complementation of said reporter enzyme;

wherein increased reporter enzyme activity in the cell compared to that which occurs in the absence of said test condition indicates increased GPCR interaction with said interacting protein partner compared to that which occurs in the absence of said test condition, and decreased reporter enzyme activity in the cell compared to that which occurs in the absence of said test condition indicates decreased GPCR interaction with interacting protein partner compared to that which occurs in the absence of said test condition.

15. The method of Claim 1, wherein said modified arrestin exhibits enhanced binding to activated, phosphorylated GPCR.

25. The method of Claim 14, wherein said modified arrestin comprises conversion of Arg170 to an amino acid selected from the group consisting of histidine, tyrosine, phenylalanine and threonine.

Cellular Expression of β_2AR-βgalΔα Fusion Protein in C2 Clones
(measured by anti-β-gal ELISA)

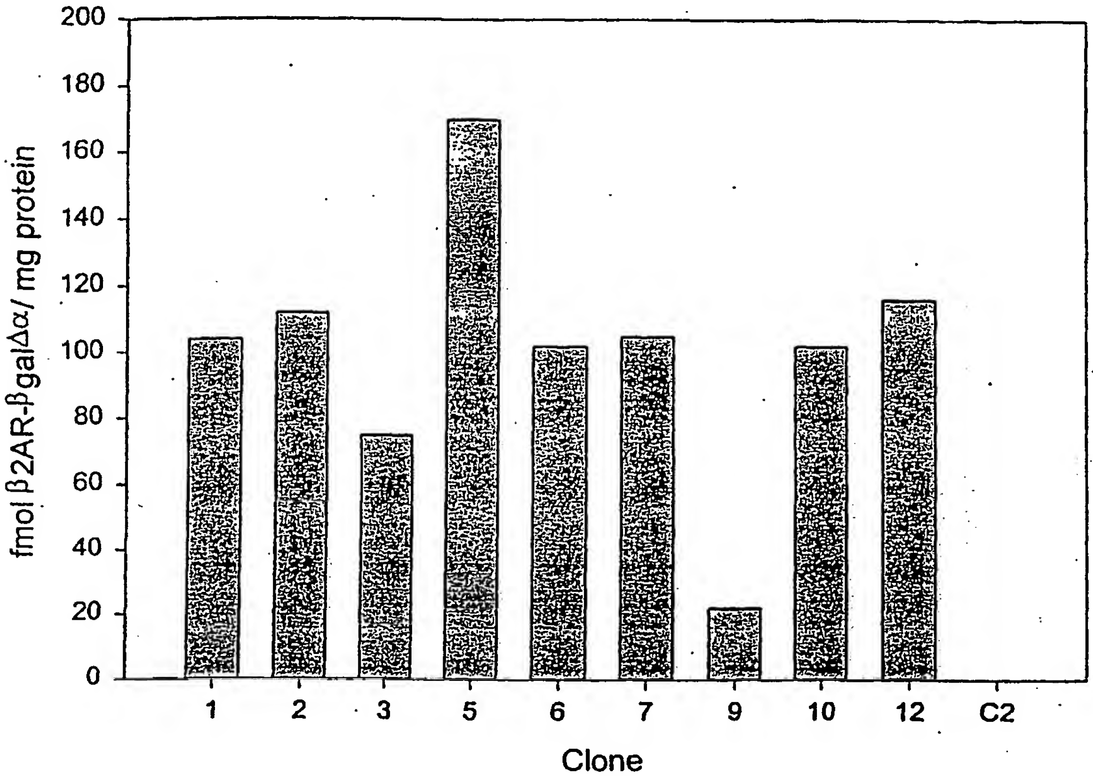


FIGURE 1A

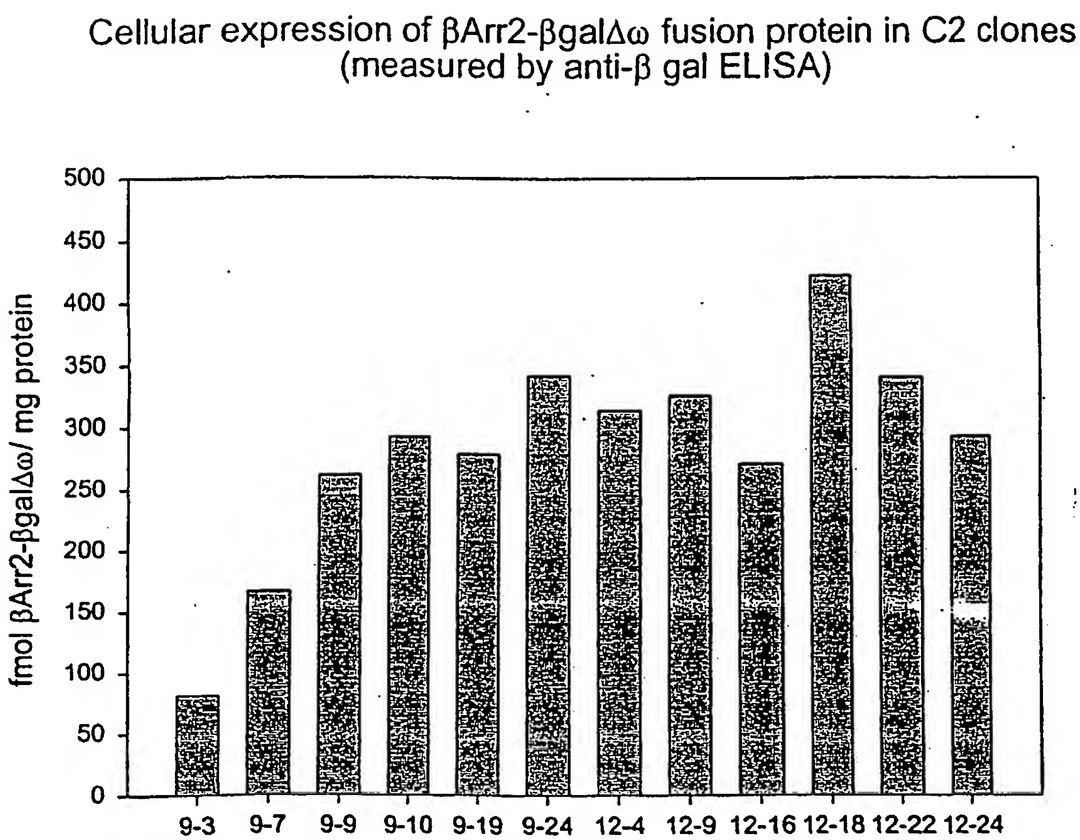

Cellular expression of βArr2-βgalΔω fusion protein in C2 clones
(measured by anti-β gal ELISA)
fmol βArr2-βgalΔω/ mg protein
500
450
400
350
300
250
200
150
100
50
0
9-3
9-7
9-9
9-10
9-19
9-24
12-4
12-9
12-16
12-18
12-22
12-24

FIGURE 1B

Agonist Stimulated cAMP Response in C2 Cells Expressing β2AR-βgalΔα

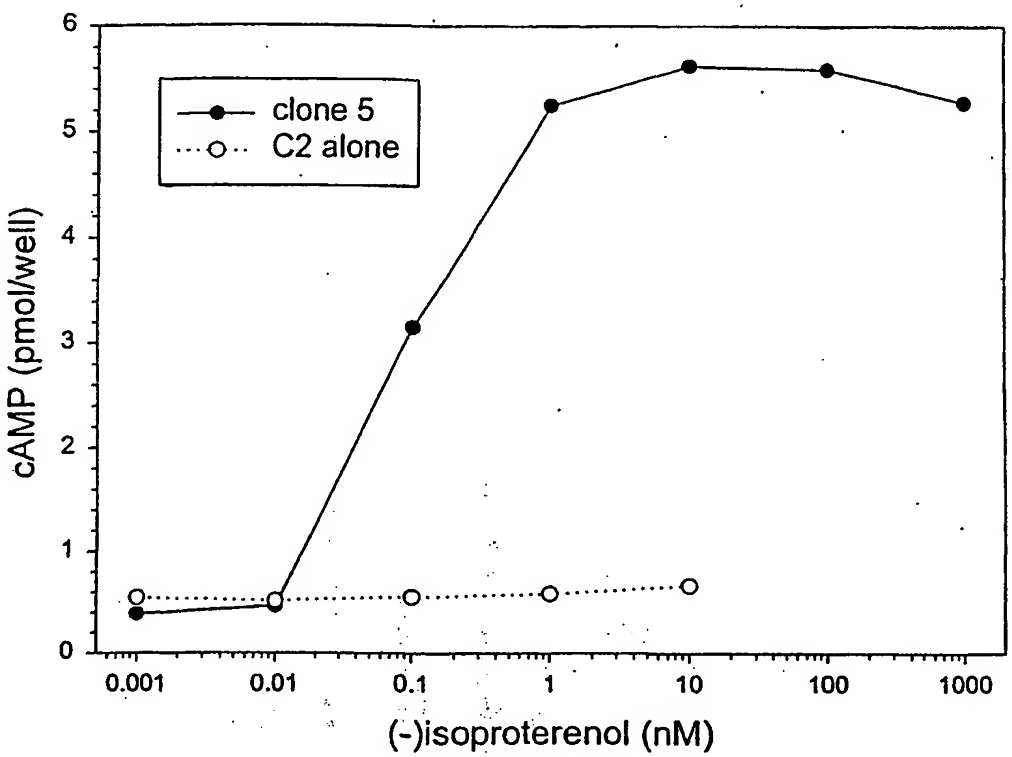


FIGURE 2

β–galactosidase Complementation as a Measurement for β2AR-βgalΔα interacting with βArrestin2-βgalΔω upon agonist Stimulation

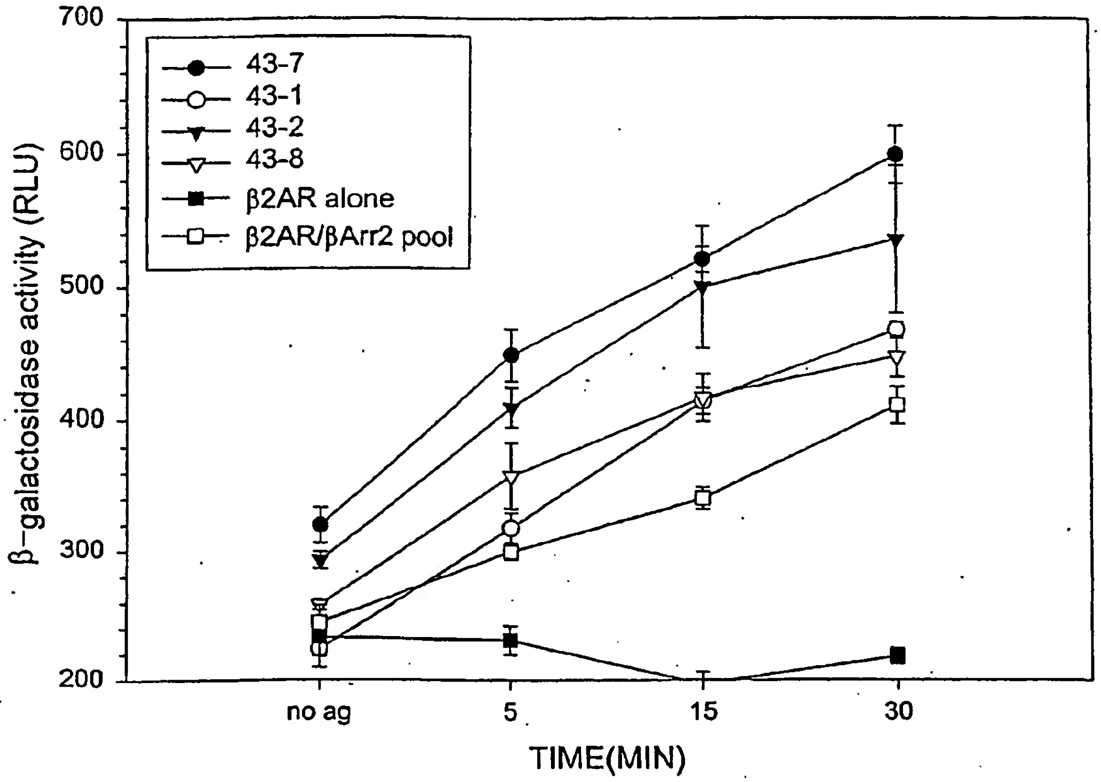


FIGURE 3A

β-galactosidase Complementation as a Measurement for β2AR-βgalΔα Interaction with βArrestin1-βgalΔω upon Agonist Stimulation

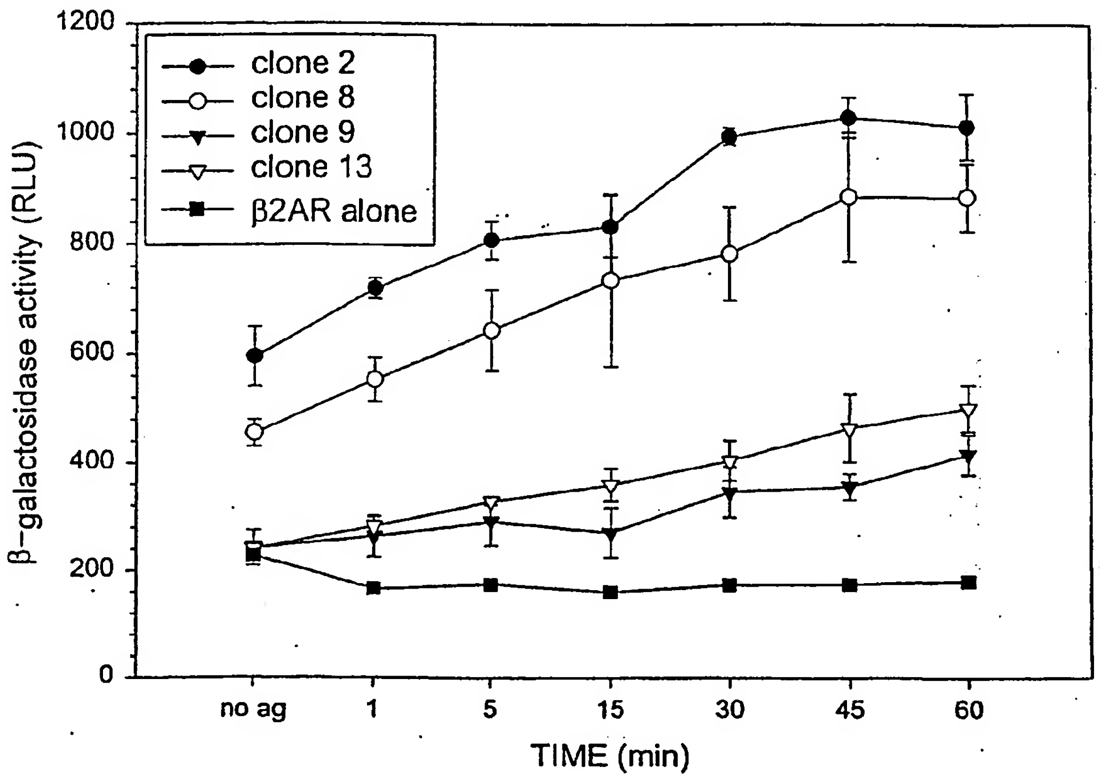


FIGURE 3B

β–galactosidase Activity in Response to Agonist in C2 Cells Coexpressing β2AR-βgalΔα and βArrestin2-βgalΔω Fusion Proteins

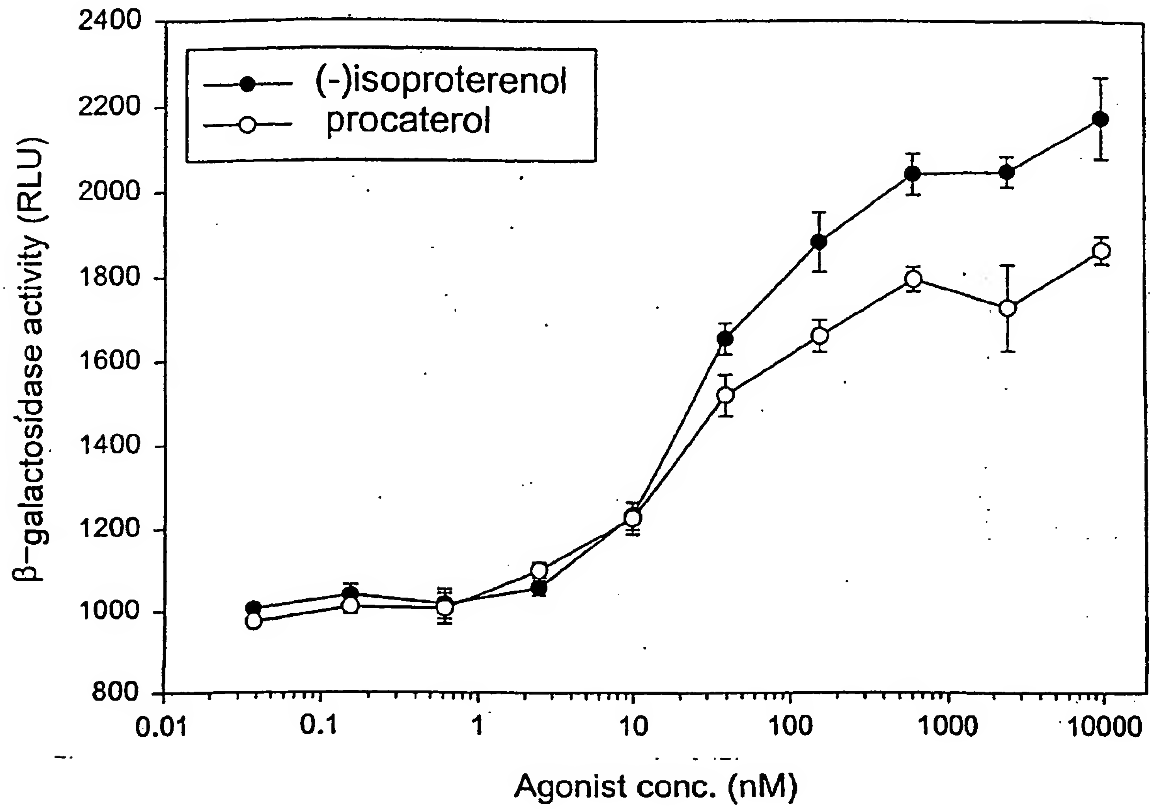


FIGURE 4A

β-galactosidase Activity in Response to Agonist in C2 Cells Coexpressing β2AR-βgalΔα and βArrestin1-βgalΔω Fusion Proteins

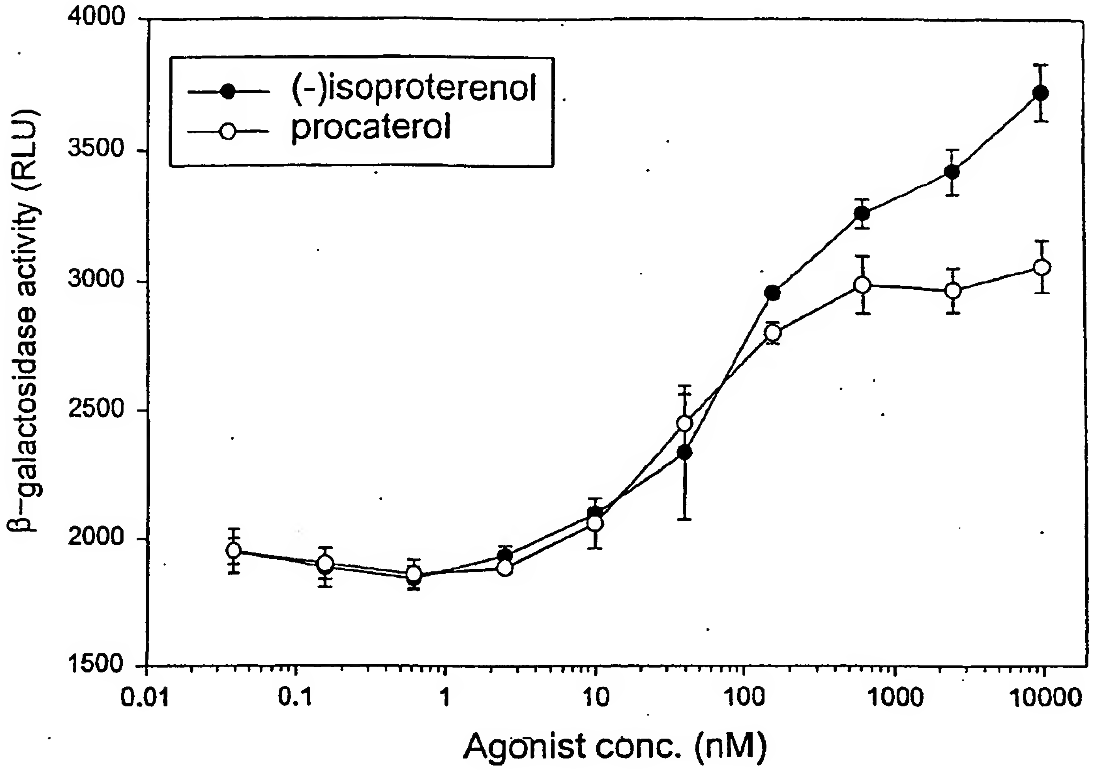


FIGURE 4B

Inhibition of β–galactosidase activity in C2 Cells Coexpressing β2AR–βgalΔα and βArrestin2-βgalΔω Fusion Proteins

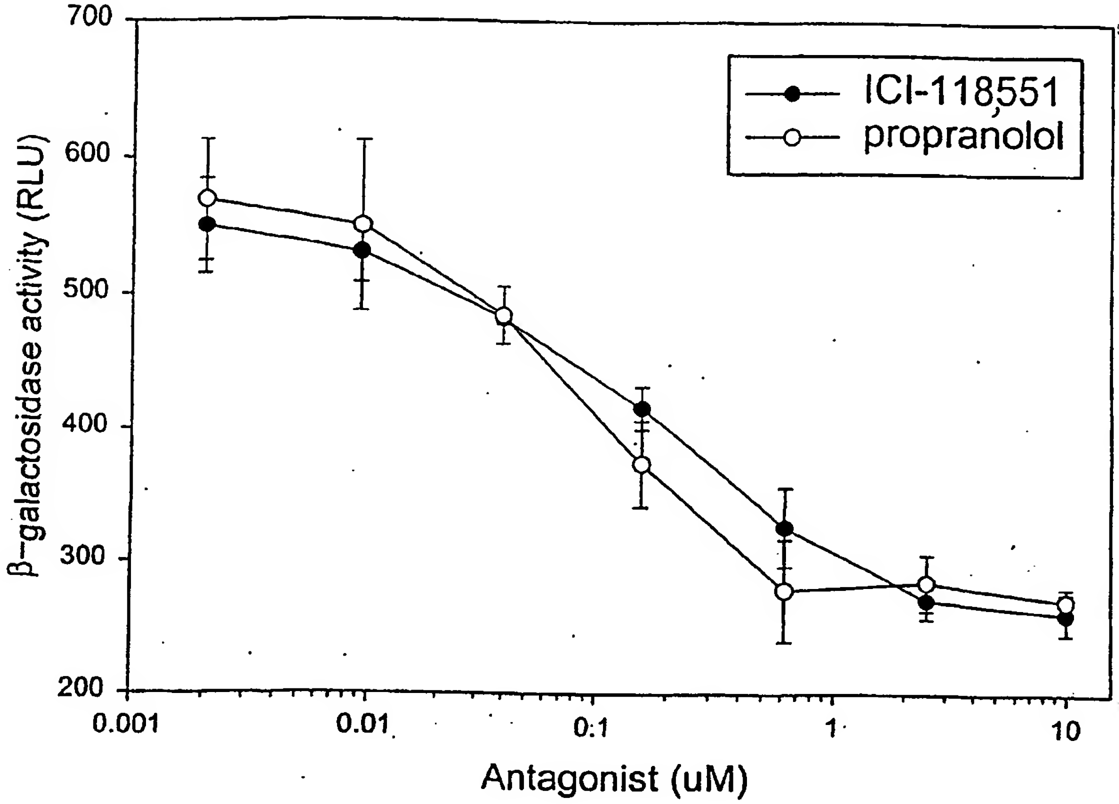


FIGURE 5A

Antagonist Inhibition of β–galactosidase Activity in C2 Cells Coexpressing β2AR-βgalΔα and βArrestin1-βgalΔω Fusion Proteins

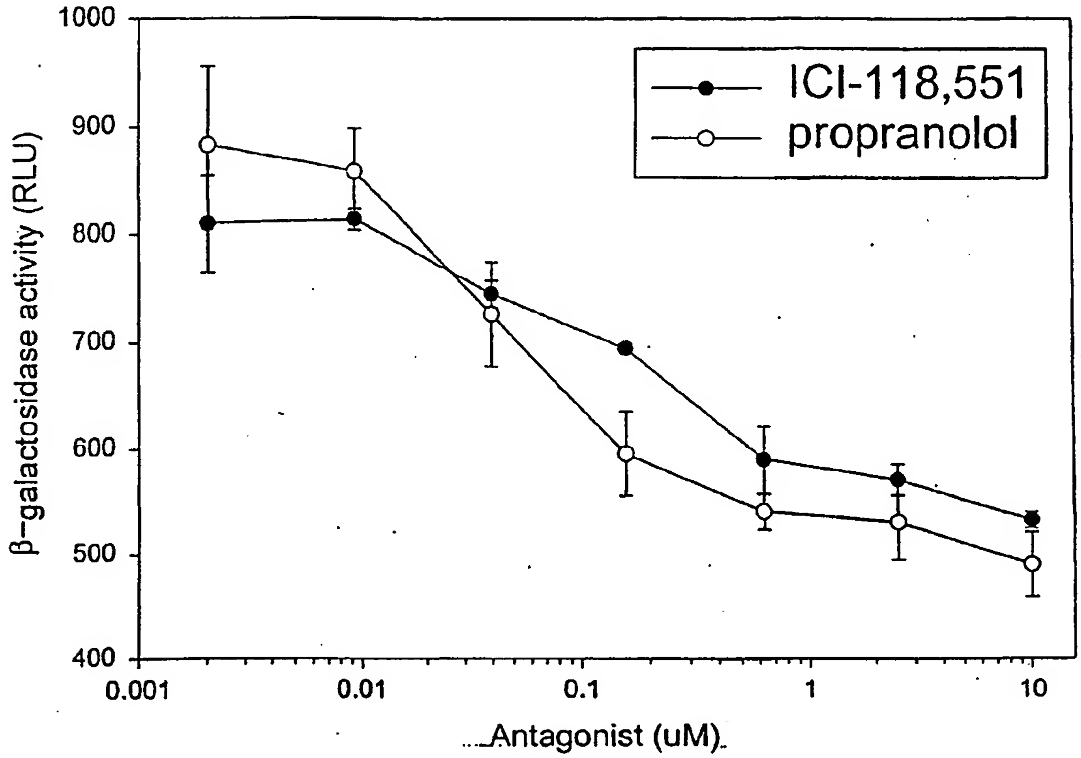


Figure 5B

Agonist Stimulated cAMP Response in Clones or Pools of C2 Cells Coexpressing A2aR-βgalΔα and βArrestin1-βgalΔω Fusion Proteins

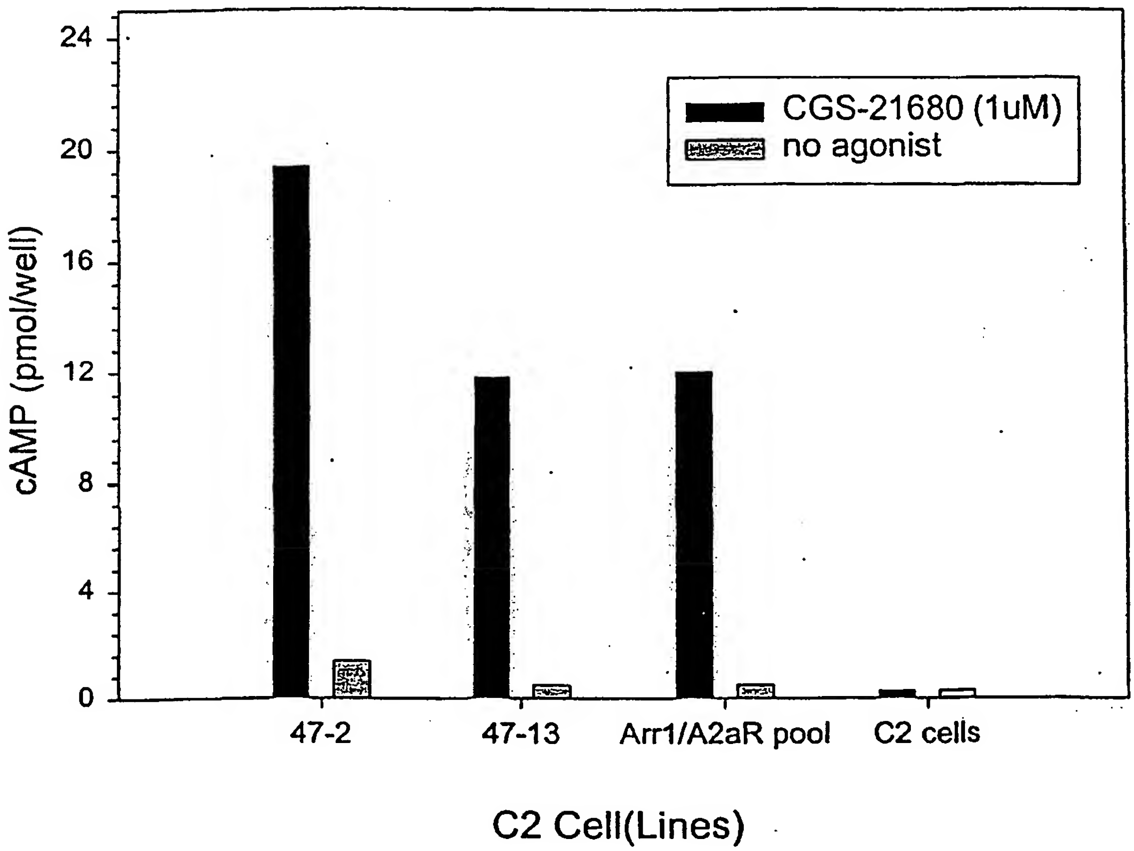


FIGURE 6

Agonist Stimulated cAMP Response in Clones or Pools of C2 Cells Expressing D1-βgalΔα and βArrestin2-βgalΔω Fusion Proteins

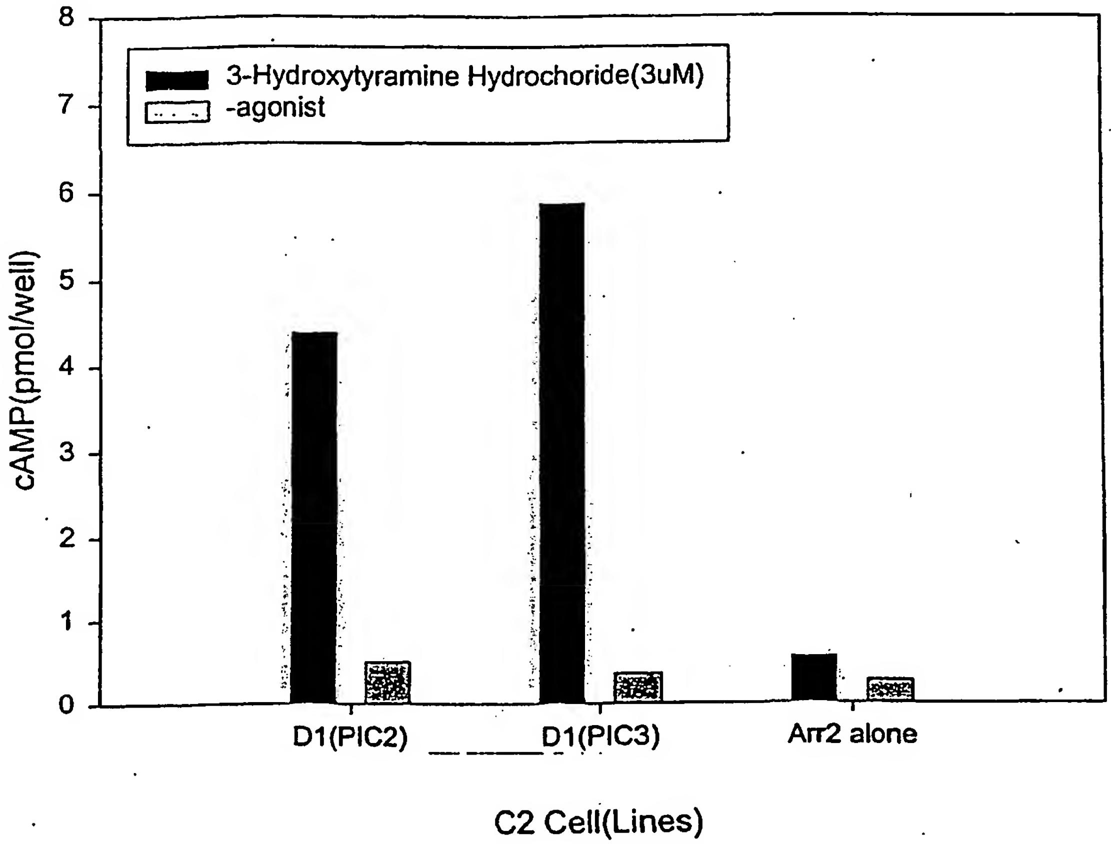


FIGURE 7

β_2AR-βgal$\Delta\omega$ and βarr2-βgal$\Delta\alpha$ Interaction in HEK293 Clones in Response to Isoproterenol Treatment (1 μM)

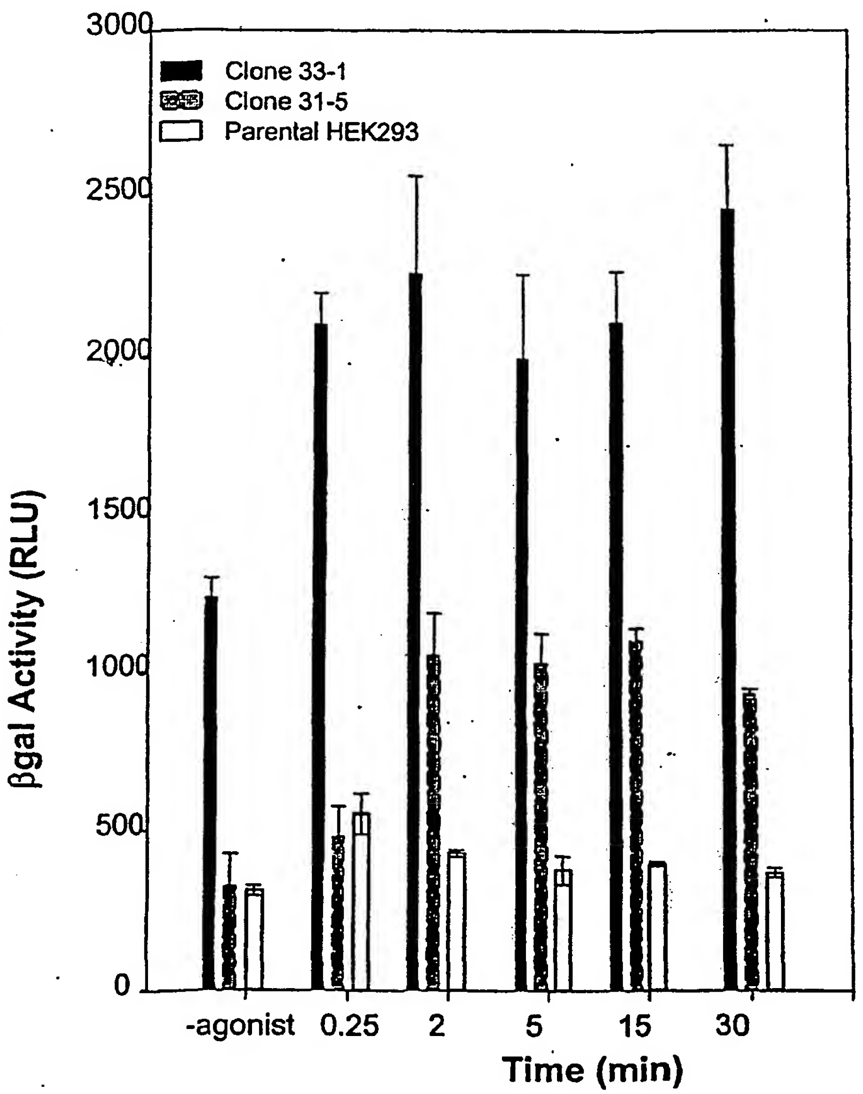


FIGURE 8A

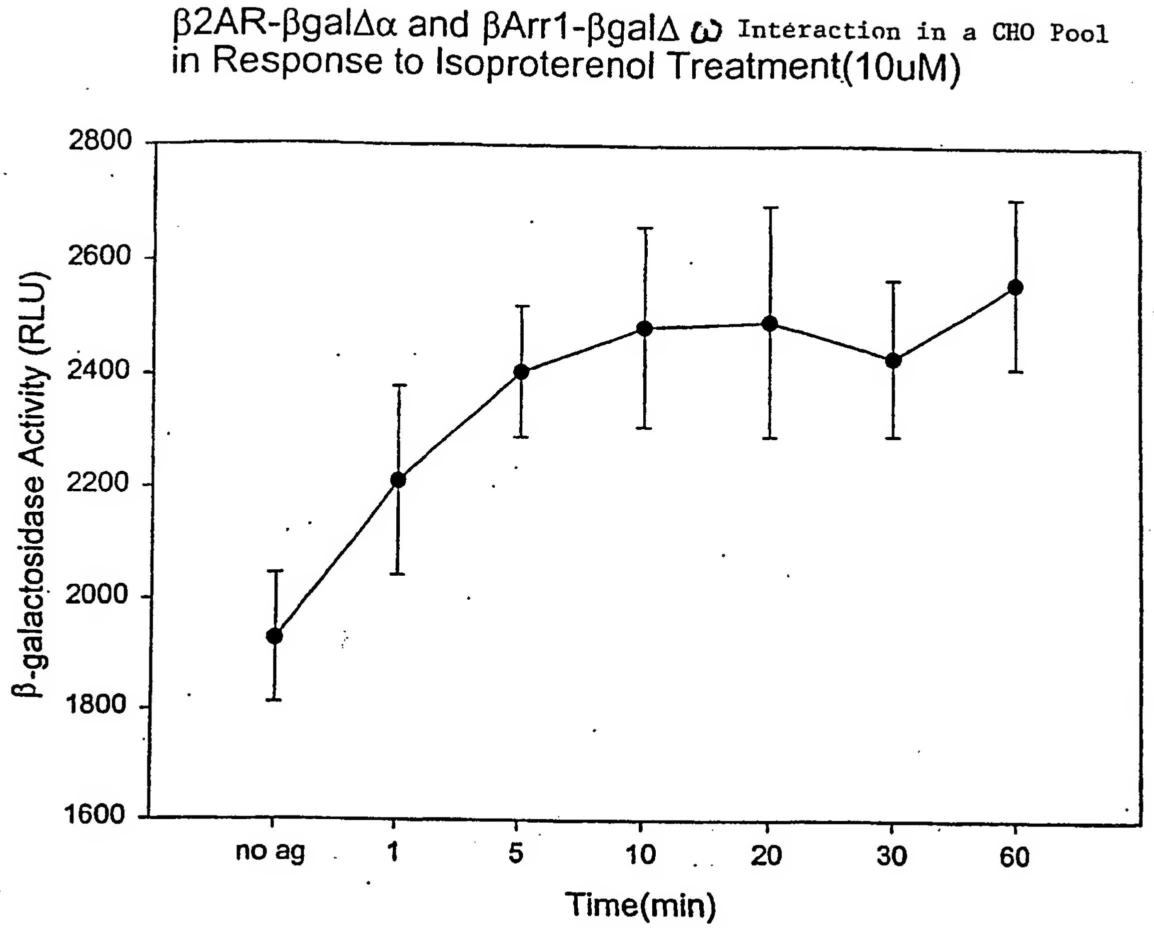

β2AR-βgalΔα and βArr1-βgalΔ ω Interaction in a CHO Pool
in Response to Isoproterenol Treatment(10uM)
2800
2600
2400
2200
2000
1800
1600
β-galactosidase Activity (RLU)
no ag
1
5
10
20
30
60
Time(min)

FIGURE 8B

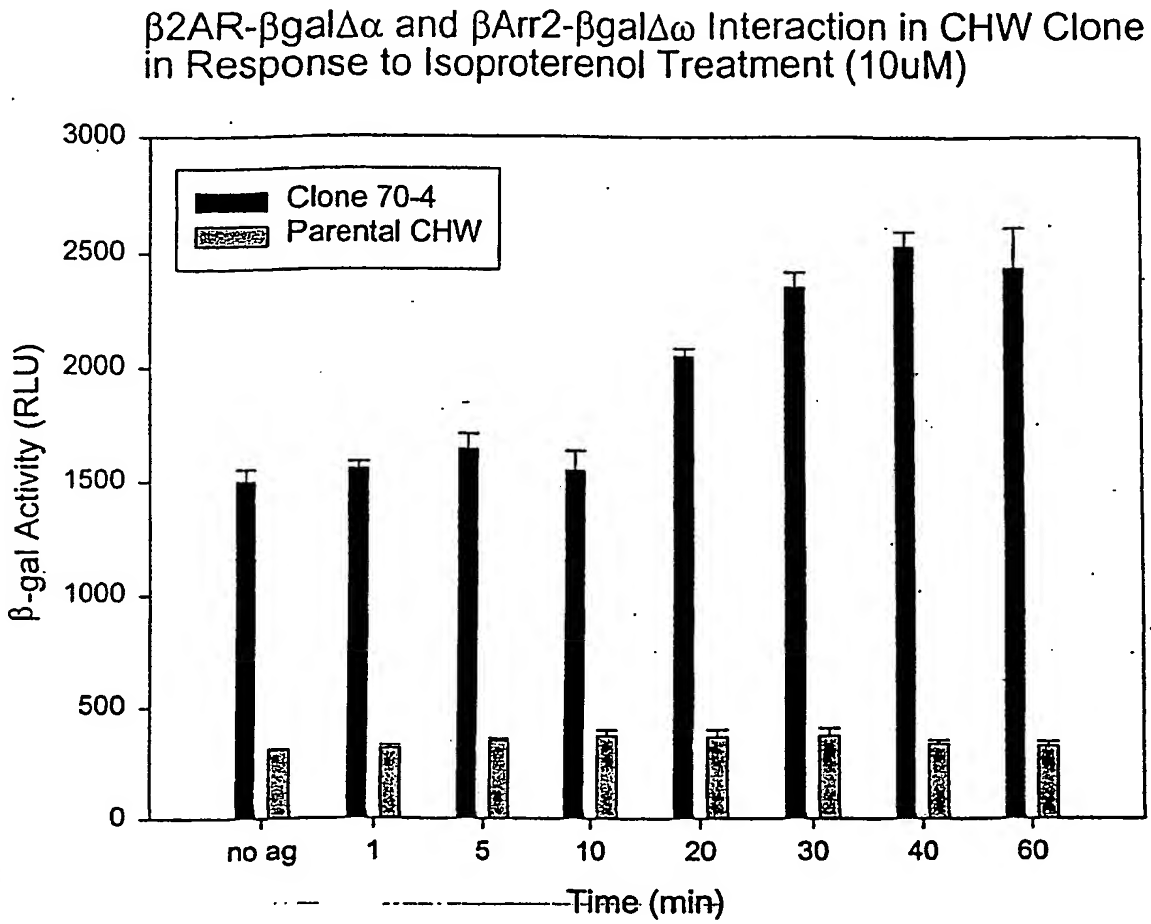

β2AR-βgalΔα and βArr2-βgalΔω Interaction in CHW Clone in Response to Isoproterenol Treatment (10uM)
3000
2500
2000
1500
1000
500
0
β-gal Activity (RLU)
Clone 70-4
Parental CHW
no ag
1
5
10
20
30
40
60
Time (min)

FIGURE 8C

β–galactosidase Complementation as a Measurement for Adrenergic Receptor Homodimerization in HEK 293 Cells Coexpressing β2AR-βgalΔα and β2AR-βgalΔω.

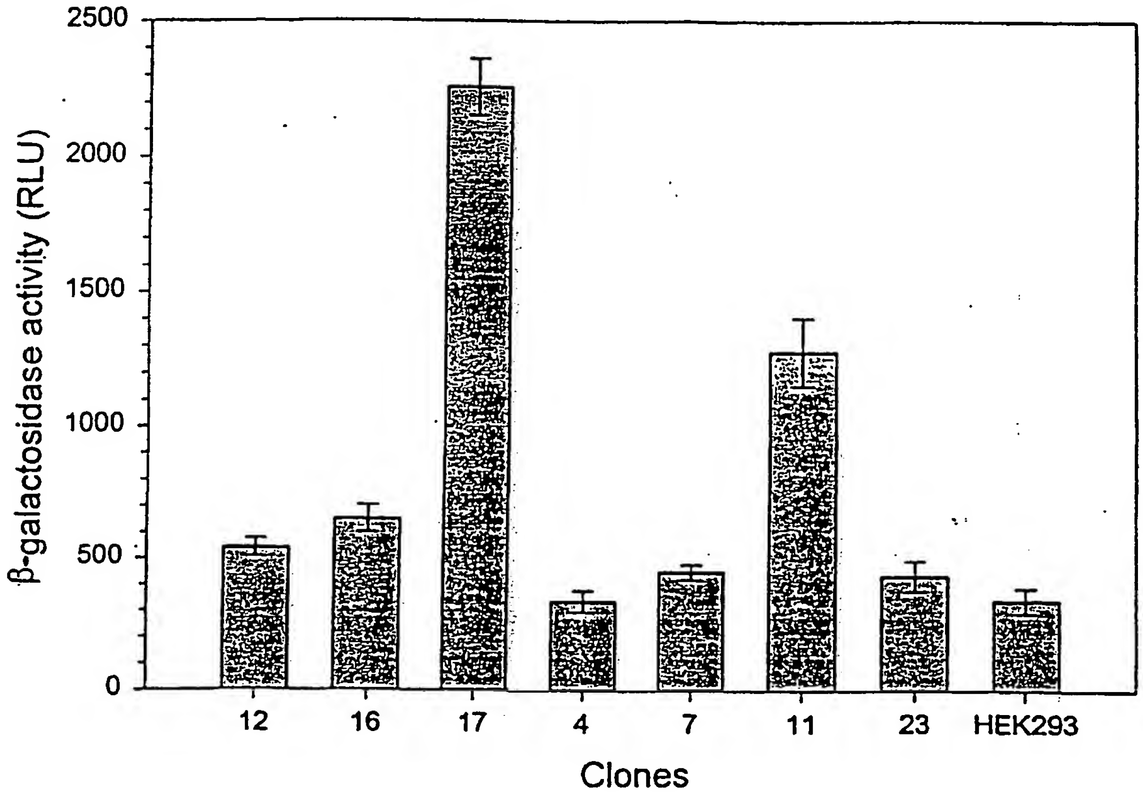


FIGURE 9A

Agonist Stimulated cAMP Response in HEK 293 Cells Coexpressing β2AR-βgalΔα and β2AR-βgalΔω

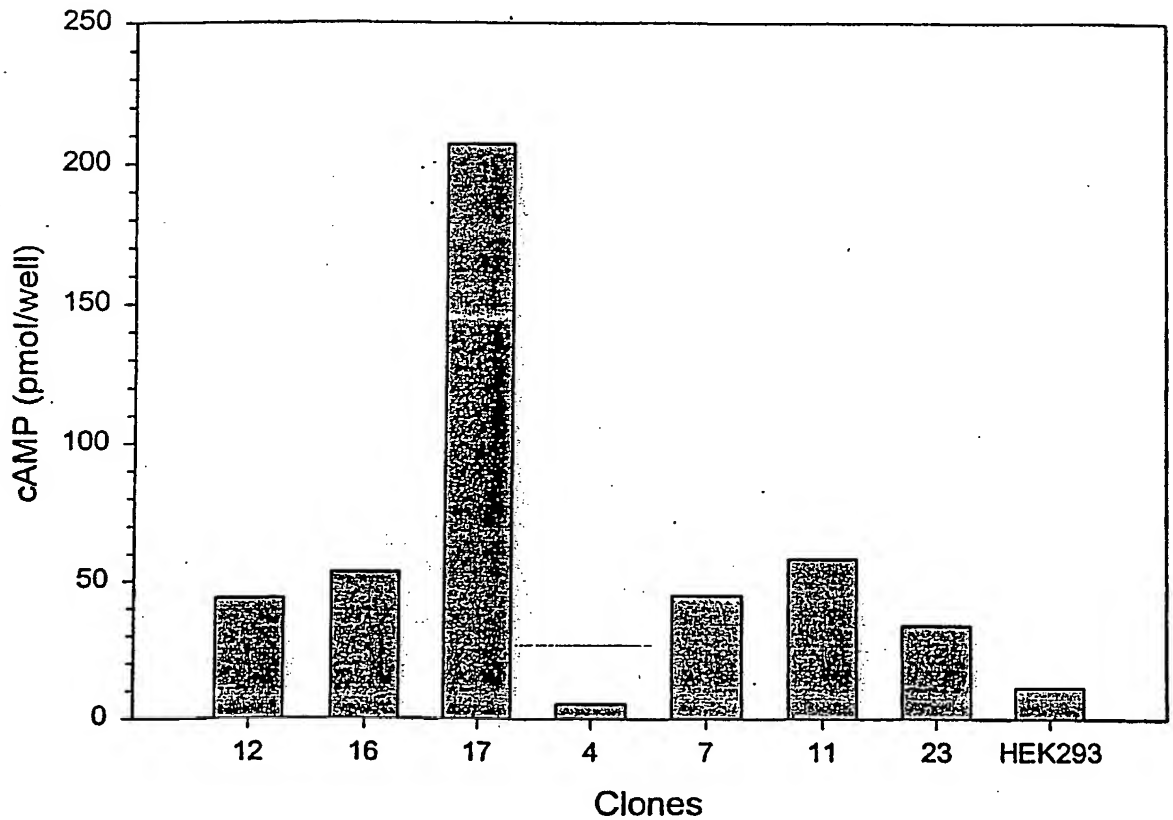


FIGURE 9B

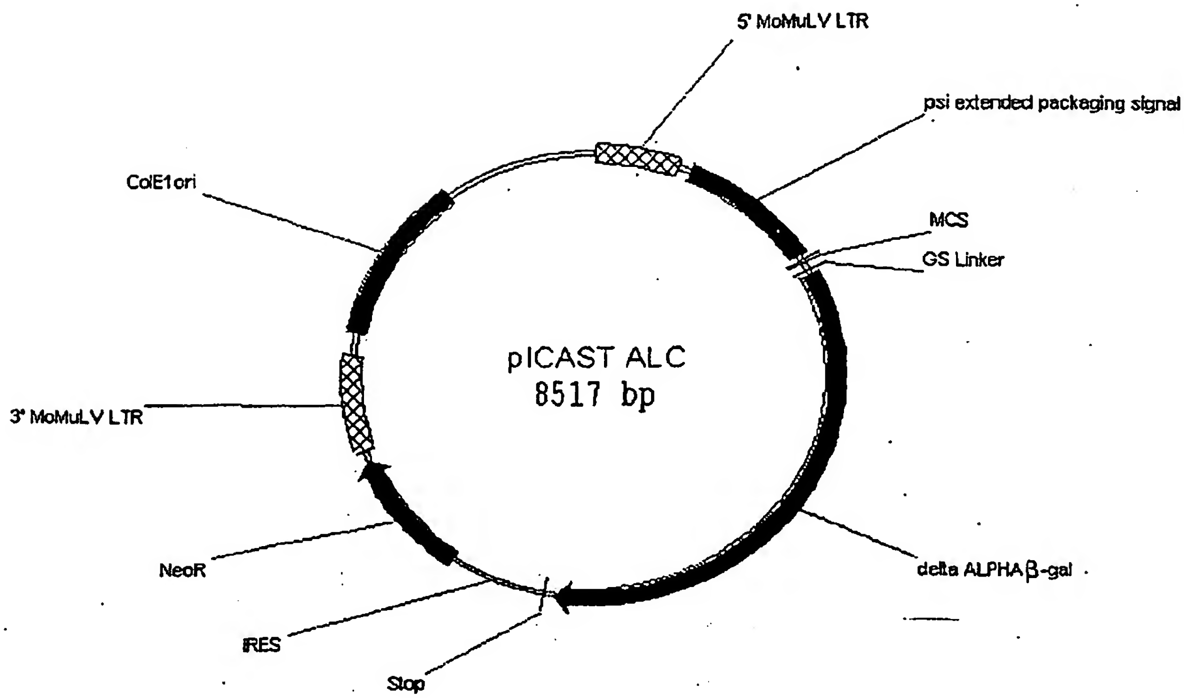

5' MoMuLV LTR
psi extended packaging signal
MCS
GS Linker
ColE1ori
pICAST ALC
8517 bp
3' MoMuLV LTR
delta ALPHA β-gal
NeoR
IRES
Stop

Figure 10A

```
1   CTGCAGCCTG AATATGGGCC AAACAGGATA TCTGTGGTAA GCAGTTCCTG
    GACGTCGGAC TTATACCCGG TTTGTCCTAT AGACACCATT CGTCAAGGAC
----------------------------------------------------------------------
51  CCCCGGCTCA GGGCCAAGAA CAGATGGAAC AGCTGAATAT GGGCCAAACA
    GGGGCCGAGT CCCGGTTCTT GTCTACCTTG TCGACTTATA CCCGGTTTGT
----------------------------------------------------------------------
101 GGATATCTGT GGTAAGCAGT TCCTGCCCCG GCTCAGGGCC AAGAACAGAT
    CCTATAGACA CCATTCGTCA AGGACGGGGC CGAGTCCCGG TTCTTGTCTA
----------------------------------------------------------------------
151 GGTCCCCAGA TGCGGTCCAG CCCTCAGCAG TTTCTAGAGA ACCATCAGAT
    CCAGGGGTCT ACGCCAGGTC GGGAGTCGTC AAAGATCTCT TGGTAGTCTA
----------------------------------------------------------------------
201 GTTTCCAGGG TGCCCCAAGG ACCTGAAATG ACCCTGTGCC TTATTTGAAC
    CAAAGGTCCC ACGGGGTTCC TGGACTTTAC TGGGACACGG AATAAACTTG
----------------------------------------------------------------------
251 TAACCAATCA GTTCGCTTCT CGCTTCTGTT CGCGCGCTTC TGCTCCCCGA
    ATTGGTTAGT CAAGCGAAGA GCGAAGACAA GCGCGCGAAG ACGAGGGGCT
----------------------------------------------------------------------
301 GCTCAATAAA AGAGCCCACA ACCCCTCACT CGGGGCGCCA GTCCTCCGAT
    CGAGTTATTT TCTCGGGTGT TGGGGAGTGA GCCCCGCGGT CAGGAGGCTA
----------------------------------------------------------------------
351 TGACTGAGTC GCCCGGGTAC CCGTGTATCC AATAAACCCT CTTGCAGTTG
    ACTGACTCAG CGGGCCCATG GGCACATAGG TTATTTGGGA GAACGTCAAC
----------------------------------------------------------------------
401 CATCCGACTT GTGGTCTCGC TGTTCCTTGG GAGGGTCTCC TCTGAGTGAT
    GTAGGCTGAA CACCAGAGCG ACAAGGAACC CTCCCAGAGG AGACTCACTA
----------------------------------------------------------------------
451 TGACTACCCG TCAGCGGGGG TCTTTCATTT GGGGGCTCGT CCGGGATCGG
    ACTGATGGGC AGTCGCCCCC AGAAAGTAAA CCCCCGAGCA GGCCCTAGCC
----------------------------------------------------------------------
501 GAGACCCCTG CCCAGGGACC ACCGACCCAC CACCGGGAGG CAAGCTGGCC
    CTCTGGGGAC GGGTCCCTGG TGGCTGGGTG GTGGCCCTCC GTTCGACCGG
----------------------------------------------------------------------
551 AGCAACTTAT CTGTGTCTGT CCGATTGTCT AGTGTCTATG ACTGATTTTA
    TCGTTGAATA GACACAGACA GGCTAACAGA TCACAGATAC TGACTAAAAT
----------------------------------------------------------------------
601 TGCGCCTGCG TCGGTACTAG TTAGCTAACT AGCTCTGTAT CTGGCGGACC
    ACGCGGACGC AGCCATGATC AATCGATTGA TCGAGACATA GACCGCCTGG
----------------------------------------------------------------------
651 CGTGGTGGAA CTGACGAGTT CTGAACACCC GGCCGCAACC CTGGGAGACG
    GCACCACCTT GACTGCTCAA GACTTGTGGG CCGGCGTTGG GACCCTCTGC
----------------------------------------------------------------------
701 TCCCAGGGAC TTTGGGGGCC GTTTTTGTGG CCCGACCTGA GGAAGGGAGT
    AGGGTCCCTG AAACCCCCGG CAAAAACACC GGGCTGGACT CCTTCCCTCA
----------------------------------------------------------------------
751 CGATGTGGAA TCCGACCCCG TCAGGATATG TGGTTCTGGT AGGAGACGAG
    GCTACACCTT AGGCTGGGGC AGTCCTATAC ACCAAGACCA TCCTCTGCTC
----------------------------------------------------------------------
801 AACCTAAAAC AGTTCCCGCC TCCGTCTGAA TTTTTGCTTT CGGTTTGGAA
    TTGGATTTTG TCAAGGGCGG AGGCAGACTT AAAAACGAAA GCCAAACCTT
----------------------------------------------------------------------
851 CCGAAGCCGC GCGTCTTGTC TGCTGCAGCA TCGTTCTGTG TTGTCTCTGT
    GGCTTCGGCG CGCAGAACAG ACGACGTCGT AGCAAGACAC AACAGAGACA
----------------------------------------------------------------------
901 CTGACTGTGT TTCTGTATTT GTCTGAAAAT TAGGGCCAGA CTGTTACCAC
    GACTGACACA AAGACATAAA CAGACTTTTA ATCCCGGTCT GACAATGGTG
----------------------------------------------------------------------
```

FIGURE 10B

```
 951  TCCCTTAAGT TTGACCTTAG GTAACTGGAA AGATGTCGAG CGGCTCGCTC
      AGGGAATTCA AACTGGAATC CATTGACCTT TCTACAGCTC GCCGAGCGAG
--------------------------------------------------------------------------------
1001  ACAACCAGTC GGTAGATGTC AAGAAGAGAC GTTGGGTTAC CTTCTGCTCT
      TGTTGGTCAG CCATCTACAG TTCTTCTCTG CAACCCAATG GAAGACGAGA
--------------------------------------------------------------------------------
1051  GCAGAATGGC CAACCTTTAA CGTCGGATGG CCGCGAGACG GCACCTTTAA
      CGTCTTACCG GTTGGAAATT GCAGCCTACC GGCGCTCTGC CGTGGAAATT
--------------------------------------------------------------------------------
1101  CCGAGACCTC ATCACCCAGG TTAAGATCAA GGTCTTTTCA CCTGGCCCGC
      GGCTCTGGAG TAGTGGGTCC AATTCTAGTT CCAGAAAAGT GGACCGGGCG
--------------------------------------------------------------------------------
1151  ATGGACACCC AGACCAGGTC CCCTACATCG TGACCTGGGA AGCCTTGGCT
      TACCTGTGGG TCTGGTCCAG GGGATGTAGC ACTGGACCCT TCGGAACCGA
--------------------------------------------------------------------------------
1201  TTTGACCCCC CTCCCTGGGT CAAGCCCTTT GTACACCCTA AGCCTCCGCC
      AAACTGGGGG GAGGGACCCA GTTCGGGAAA CATGTGGGAT TCGGAGGCGG
--------------------------------------------------------------------------------
1251  TCCTCTTCCT CCATCCGCCC CGTCTCTCCC CCTTGAACCT CCTCGTTCGA
      AGGAGAAGGA GGTAGGCGGG GCAGAGAGGG GGAACTTGGA GGAGCAAGCT
--------------------------------------------------------------------------------
1301  CCCCGCCTCG ATCCTCCCTT TATCCAGCCC TCACTCCTTC TCTAGGCGCC
      GGGGCGGAGC TAGGAGGGAA ATAGGTCGGG AGTGAGGAAG AGATCCGCGG
--------------------------------------------------------------------------------
1351  GGCCGCTCTA GCCCATTAAT ACGACTCACT ATAGGGCGAT TCGAATCAGG
      CCGGCGAGAT CGGGTAATTA TGCTGAGTGA TATCCCGCTA AGCTTAGTCC
--------------------------------------------------------------------------------
1401  CCTTGGCGCG CCGGATCCTT AATTAAGCGC AATTGGGAGG TGGCGGTAGC
      GGAACCGCGC GGCCTAGGAA TTAATTCGCG TTAACCCTCC ACCGCCATCG
--------------------------------------------------------------------------------
  +2          M  G   V  I  T   D  S  L   A  V  V   A  R  T  D
             ]--------------------------------------------------
1451  CTCGAGATGG GCGTGATTAC GGATTCACTG GCCGTCGTGG CCCGCACCGA
      GAGCTCTACC CGCACTAATG CCTAAGTGAC CGGCAGCACC GGGCGTGGCT
--------------------------------------------------------------------------------
  +2    R  P  S   Q  Q  L  R   S  L  N   G  E  W   R  F  A
      --------------------------------------------------
1501  TCGCCCTTCC CAACAGTTAC GCAGCCTGAA TGGCGAATGG CGCTTTGCCT
      AGCGGGAAGG GTTGTCAATG CGTCGGACTT ACCGCTTACC GCGAAACGGA
--------------------------------------------------------------------------------
  +2  W  F  P  A   P  E  A   V  P  E  S   W  L  E   C  D  L
      --------------------------------------------------
1551  GGTTTCCGGC ACCAGAAGCG GTGCCGGAAA GCTGGCTGGA GTGCGATCTT
      CCAAAGGCCG TGGTCTTCGC CACGGCCTTT CGACCGACCT CACGCTAGAA
--------------------------------------------------------------------------------
  +2   P  E  A  D   T  V  V   V  P  S   N  W  Q  M   H  G  Y
      --------------------------------------------------
1601  CCTGAGGCCG ATACTGTCGT CGTCCCCTCA AACTGGCAGA TGCACGGTTA
      GGACTCCGGC TATGACAGCA GCAGGGGAGT TTGACCGTCT ACGTGCCAAT
--------------------------------------------------------------------------------
  +2    D  A  P   I  Y  T  N   V  T  Y   P  I  T   V  N  P
      --------------------------------------------------
1651  CGATGCGCCC ATCTACACCA ACGTGACCTA TCCCATTACG GTCAATCCGC
      GCTACGCGGG TAGATGTGGT TGCACTGGAT AGGGTAATGC CAGTTAGGCG
--------------------------------------------------------------------------------
```

```
+2 P  F  V  P   T  E  N   P  T  G  C   Y  S  L   T  F  N
   ---------------------------------------------------
1701  CGTTTGTTCC CACGGAGAAT CCGACGGGTT GTTACTCGCT CACATTTAAT
      GCAAACAAGG GTGCCTCTTA GGCTGCCCAA CAATGAGCGA GTGTAAATTA
---------------------------------------------------------------------------
+2  V  D  E  S   W  L  Q   E  G  Q   T  R  I  I   F  D  G
   ---------------------------------------------------
1751  GTTGATGAAA GCTGGCTACA GGAAGGCCAG ACGCGAATTA TTTTTGATGG
      CAACTACTTT CGACCGATGT CCTTCCGGTC TGCGCTTAAT AAAAACTACC
---------------------------------------------------------------------------
+2  V  N  S   A  F  H  L   W  C  N   G  R  W   V  G  Y
   ---------------------------------------------------
1801  CGTTAACTCG GCGTTTCATC TGTGGTGCAA CGGGCGCTGG GTCGGTTACG
      GCAATTGAGC CGCAAAGTAG ACACCACGTT GCCCGCGACC CAGCCAATGC
---------------------------------------------------------------------------
+2 G  Q  D  S   R  L  P   S  E  F  D   L  S  A   F  L  R
   ---------------------------------------------------
1851  GCCAGGACAG TCGTTTGCCG TCTGAATTTG ACCTGAGCGC ATTTTTACGC
      CGGTCCTGTC AGCAAACGGC AGACTTAAAC TGGACTCGCG TAAAAATGCG
---------------------------------------------------------------------------
+2  A  G  E  N   R  L  A   V  M  V   L  R  W  S   D  G  S
   ---------------------------------------------------
1901  GCCGGAGAAA ACCGCCTCGC GGTGATGGTG CTGCGCTGGA GTGACGGCAG
      CGGCCTCTTT TGGCGGAGCG CCACTACCAC GACGCGACCT CACTGCCGTC
---------------------------------------------------------------------------
+2   Y  L  E   D  Q  D  M   W  R  M   S  G  I   F  R  D
   ---------------------------------------------------
1951  TTATCTGGAA GATCAGGATA TGTGGCGGAT GAGCGGCATT TTCCGTGACG
      AATAGACCTT CTAGTCCTAT ACACCGCCTA CTCGCCGTAA AAGGCACTGC
---------------------------------------------------------------------------
+2 V  S  L  L   H  K  P   T  T  Q  I   S  D  F   H  V  A
   ---------------------------------------------------
2001  TCTCGTTGCT GCATAAACCG ACTACACAAA TCAGCGATTT CCATGTTGCC
      AGAGCAACGA CGTATTTGGC TGATGTGTTT AGTCGCTAAA GGTACAACGG
---------------------------------------------------------------------------
+2  T  R  F  N   D  D  F   S  R  A   V  L  E  A   E  V  Q
   ---------------------------------------------------
2051  ACTCGCTTTA ATGATGATTT CAGCCGCGCT GTACTGGAGG CTGAAGTTCA
      TGAGCGAAAT TACTACTAAA GTCGGCGCGA CATGACCTCC GACTTCAAGT
---------------------------------------------------------------------------
+2   M  C  G   E  L  R  D   Y  L  R   V  T  V   S  L  W
   ---------------------------------------------------
2101  GATGTGCGGC GAGTTGCGTG ACTACCTACG GGTAACAGTT TCTTTATGGC
      CTACACGCCG CTCAACGCAC TGATGGATGC CCATTGTCAA AGAAATACCG
---------------------------------------------------------------------------
+2 Q  G  E  T   Q  V  A   S  G  T  A   P  F  G   G  E  I
   ---------------------------------------------------
2151  AGGGTGAAAC GCAGGTCGCC AGCGGCACCG CGCCTTTCGG CGGTGAAATT
      TCCCACTTTG CGTCCAGCGG TCGCCGTGGC GCGGAAAGCC GCCACTTTAA
---------------------------------------------------------------------------
+2  I  D  E  R   G  G  Y   A  D  R   V  T  L  R   L  N  V
   ---------------------------------------------------
2201  ATCGATGAGC GTGGTGGTTA TGCCGATCGC GTCACACTAC GTCTGAACGT
      TAGCTACTCG CACCACCAAT ACGGCTAGCG CAGTGTGATG CAGACTTGCA
---------------------------------------------------------------------------
+2   E  N  P   K  L  W  S   A  E  I   P  N  L   Y  R  A
   ---------------------------------------------------
2251  CGAAAACCCG AAACTGTGGA GCGCCGAAAT CCCGAATCTC TATCGTGCGG
      GCTTTTGGGC TTTGACACCT CGCGGCTTTA GGGCTTAGAG ATAGCACGCC
---------------------------------------------------------------------------
```

```
+2 V  V  E  L  H  T  A  D  G  T  L  I  E  A  E  A  C
   ---------------------------------------------------
2301 TGGTTGAACT GCACACCGCC GACGGCACGC TGATTGAAGC AGAAGCCTGC
     ACCAACTTGA CGTGTGGCGG CTGCCGTGCG ACTAACTTCG TCTTCGGACG
-----------------------------------------------------------
+2 D  V  G  F  R  E  V  R  I  E  N  G  L  L  L  L  N
   ---------------------------------------------------
2351 GATGTCGGTT TCCGCGAGGT GCGGATTGAA AATGGTCTGC TGCTGCTGAA
     CTACAGCCAA AGGCGCTCCA CGCCTAACTT TTACCAGACG ACGACGACTT
-----------------------------------------------------------
+2 G  K  P  L  L  I  R  G  V  N  R  H  E  H  H  P
   ---------------------------------------------------
2401 CGGCAAGCCG TTGCTGATTC GAGGCGTTAA CCGTCACGAG CATCATCCTC
     GCCGTTCGGC AACGACTAAG CTCCGCAATT GGCAGTGCTC GTAGTAGGAG
-----------------------------------------------------------
+2 L  H  G  Q  V  M  D  E  Q  T  M  V  Q  D  I  L  L
   ---------------------------------------------------
2451 TGCATGGTCA GGTCATGGAT GAGCAGACGA TGGTGCAGGA TATCCTGCTG
     ACGTACCAGT CCAGTACCTA CTCGTCTGCT ACCACGTCCT ATAGGACGAC
-----------------------------------------------------------
+2 M  K  Q  N  N  F  N  A  V  R  C  S  H  Y  P  N  H
   ---------------------------------------------------
2501 ATGAAGCAGA ACAACTTTAA CGCCGTGCGC TGTTCGCATT ATCCGAACCA
     TACTTCGTCT TGTTGAAATT GCGGCACGCG ACAAGCGTAA TAGGCTTGGT
-----------------------------------------------------------
+2 P  L  W  Y  T  L  C  D  R  Y  G  L  Y  V  V  D
   ---------------------------------------------------
2551 TCCGCTGTGG TACACGCTGT GCGACCGCTA CGGCCTGTAT GTGGTGGATG
     AGGCGACACC ATGTGCGACA CGCTGGCGAT GCCGGACATA CACCACCTAC
-----------------------------------------------------------
+2 E  A  N  I  E  T  H  G  M  V  P  M  N  R  L  T  D
   ---------------------------------------------------
2601 AAGCCAATAT TGAAACCCAC GGCATGGTGC CAATGAATCG TCTGACCGAT
     TTCGGTTATA ACTTTGGGTG CCGTACCACG GTTACTTAGC AGACTGGCTA
-----------------------------------------------------------
+2 D  P  R  W  L  P  A  M  S  E  R  V  T  R  M  V  Q
   ---------------------------------------------------
2651 GATCCGCGCT GGCTACCGGC GATGAGCGAA CGCGTAACGC GAATGGTGCA
     CTAGGCGCGA CCGATGGCCG CTACTCGCTT GCGCATTGCG CTTACCACGT
-----------------------------------------------------------
+2 R  D  R  N  H  P  S  V  I  I  W  S  L  G  N  E
   ---------------------------------------------------
2701 GCGCGATCGT AATCACCCGA GTGTGATCAT CTGGTCGCTG GGGAATGAAT
     CGCGCTAGCA TTAGTGGGCT CACACTAGTA GACCAGCGAC CCCTTACTTA
-----------------------------------------------------------
+2 S  G  H  G  A  N  H  D  A  L  Y  R  W  I  K  S  V
   ---------------------------------------------------
2751 CAGGCCACGG CGCTAATCAC GACGCGCTGT ATCGCTGGAT CAAATCTGTC
     GTCCGGTGCC GCGATTAGTG CTGCGCGACA TAGCGACCTA GTTTAGACAG
-----------------------------------------------------------
+2 D  P  S  R  P  V  Q  Y  E  G  G  G  A  D  T  T  A
   ---------------------------------------------------
2801 GATCCTTCCC GCCCGGTGCA GTATGAAGGC GGCGGAGCCG ACACCACGGC
     CTAGGAAGGG CGGGCCACGT CATACTTCCG CCGCCTCGGC TGTGGTGCCG
-----------------------------------------------------------
+2 T  D  I  I  C  P  M  Y  A  R  V  D  E  D  Q  P
   ---------------------------------------------------
2851 CACCGATATT ATTTGCCCGA TGTACGCGCG CGTGGATGAA GACCAGCCCT
     GTGGCTATAA TAAACGGGCT ACATGCGCGC GCACCTACTT CTGGTCGGGA
-----------------------------------------------------------
```

```
+2 F  P  A  V   P  K  W   S  I  K  K   W  L  S   L  P  G
   ----------------------------------------------------
2901  TCCCGGCTGT GCCGAAATGG TCCATCAAAA AATGGCTTTC GCTACCTGGA
      AGGGCCGACA CGGCTTTACC AGGTAGTTTT TTACCGAAAG CGATGGACCT
---------------------------------------------------------------------------
+2  E  T  R  P   L  I  L   C  E  Y   A  H  A  M   G  N  S
   ----------------------------------------------------
2951  GAGACGCGCC CGCTGATCCT TTGCGAATAC GCCCACGCGA TGGGTAACAG
      CTCTGCGCGG GCGACTAGGA AACGCTTATG CGGGTGCGCT ACCCATTGTC
---------------------------------------------------------------------------
+2  L  G  G   F  A  K  Y   W  Q  A   F  R  Q   Y  P  R
   ----------------------------------------------------
3001  TCTTGGCGGT TTCGCTAAAT ACTGGCAGGC GTTTCGTCAG TATCCCCGTT
      AGAACCGCCA AAGCGATTTA TGACCGTCCG CAAAGCAGTC ATAGGGGCAA
---------------------------------------------------------------------------
+2 L  Q  G  G   F  V  W   D  W  V  D   Q  S  L   I  K  Y
   ----------------------------------------------------
3051  TACAGGGCGG CTTCGTCTGG GACTGGGTGG ATCAGTCGCT GATTAAATAT
      ATGTCCCGCC GAAGCAGACC CTGACCCACC TAGTCAGCGA CTAATTTATA
---------------------------------------------------------------------------
+2  D  E  N  G   N  P  W   S  A  Y   G  G  D  F   G  D  T
   ----------------------------------------------------
3101  GATGAAAACG GCAACCCGTG GTCGGCTTAC GGCGGTGATT TTGGCGATAC
      CTACTTTTGC CGTTGGGCAC CAGCCGAATG CCGCCACTAA AACCGCTATG
---------------------------------------------------------------------------
+2   P  N  D   R  Q  F  C   M  N  G   L  V  F   A  D  R
   ----------------------------------------------------
3151  GCCGAACGAT CGCCAGTTCT GTATGAACGG TCTGGTCTTT GCCGACCGCA
      CGGCTTGCTA GCGGTCAAGA CATACTTGCC AGACCAGAAA CGGCTGGCGT
---------------------------------------------------------------------------
+2 T  P  H  P   A  L  T   E  A  K  H   Q  Q  Q   F  F  Q
   ----------------------------------------------------
3201  CGCCGCATCC AGCGCTGACG GAAGCAAAAC ACCAGCAGCA GTTTTTCCAG
      GCGGCGTAGG TCGCGACTGC CTTCGTTTTG TGGTCGTCGT CAAAAAGGTC
---------------------------------------------------------------------------
+2  F  R  L  S   G  Q  T   I  E  V   T  S  E  Y   L  F  R
   ----------------------------------------------------
3251  TTCCGTTTAT CCGGGCAAAC CATCGAAGTG ACCAGCGAAT ACCTGTTCCG
      AAGGCAAATA GGCCCGTTTG GTAGCTTCAC TGGTCGCTTA TGGACAAGGC
---------------------------------------------------------------------------
+2   H  S  D   N  E  L  L   H  W  M   V  A  L   D  G  K
   ----------------------------------------------------
3301  TCATAGCGAT AACGAGCTCC TGCACTGGAT GGTGGCGCTG GATGGTAAGC
      AGTATCGCTA TTGCTCGAGG ACGTGACCTA CCACCGCGAC CTACCATTCG
---------------------------------------------------------------------------
+2 P  L  A  S   G  E  V   P  L  D  V   A  P  Q   G  K  Q
   ----------------------------------------------------
3351  CGCTGGCAAG CGGTGAAGTG CCTCTGGATG TCGCTCCACA AGGTAAACAG
      GCGACCGTTC GCCACTTCAC GGAGACCTAC AGCGAGGTGT TCCATTTGTC
---------------------------------------------------------------------------
+2  L  I  E  L   P  E  L   P  Q  P   E  S  A  G   Q  L  W
   ----------------------------------------------------
3401  TTGATTGAAC TGCCTGAACT ACCGCAGCCG GAGAGCGCCG GGCAACTCTG
      AACTAACTTG ACGGACTTGA TGGCGTCGGC CTCTCGCGGC CCGTTGAGAC
---------------------------------------------------------------------------
+2   L  T  V   R  V  V  Q   P  N  A   T  A  W   S  E  A
   ----------------------------------------------------
3451  GCTCACAGTA CGCGTAGTGC AACCGAACGC GACCGCATGG TCAGAAGCCG
      CGAGTGTCAT GCGCATCACG TTGGCTTGCG CTGGCGTACC AGTCTTCGGC
---------------------------------------------------------------------------
```

```
+2 G  H  I  S   A  W  Q   Q  W  R  L   A  E  N   L  S  V
     ----------------------------------------------------
3501 GGCACATCAG CGCCTGGCAG CAGTGGCGTC TGGCGGAAAA CCTCAGTGTG
     CCGTGTAGTC GCGGACCGTC GTCACCGCAG ACCGCCTTTT GGAGTCACAC
-------------------------------------------------------------
+2   T  L  P  A   A  S  H   A  I  P   H  L  T  T   S  E  M
     ----------------------------------------------------
3551 ACGCTCCCCG CCGCGTCCCA CGCCATCCCG CATCTGACCA CCAGCGAAAT
     TGCGAGGGGC GGCGCAGGGT GCGGTAGGGC GTAGACTGGT GGTCGCTTTA
-------------------------------------------------------------
+2   D  F  C   I  E  L  G   N  K  R   W  Q  F   N  R  Q
     ----------------------------------------------------
3601 GGATTTTTGC ATCGAGCTGG GTAATAAGCG TTGGCAATTT AACCGCCAGT
     CCTAAAAACG TAGCTCGACC CATTATTCGC AACCGTTAAA TTGGCGGTCA
-------------------------------------------------------------
+2 S  G  F  L   S  Q  M   W  I  G  D   K  K  Q  L  L  T
     ----------------------------------------------------
3651 CAGGCTTTCT TTCACAGATG TGGATTGGCG ATAAAAAACA ACTGCTGACG
     GTCCGAAAGA AAGTGTCTAC ACCTAACCGC TATTTTTTGT TGACGACTGC
-------------------------------------------------------------
+2   P  L  R  D   Q  F  T   R  A  P   L  D  N  D   I  G  V
     ----------------------------------------------------
3701 CCGCTGCGCG ATCAGTTCAC CCGTGCACCG CTGGATAACG ACATTGGCGT
     GGCGACGCGC TAGTCAAGTG GGCACGTGGC GACCTATTGC TGTAACCGCA
-------------------------------------------------------------
+2   S  E  A   T  R  I  D   P  N  A   W  V  E   R  W  K
     ----------------------------------------------------
3751 AAGTGAAGCG ACCCGCATTG ACCCTAACGC CTGGGTCGAA CGCTGGAAGG
     TTCACTTCGC TGGGCGTAAC TGGGATTGCG GACCCAGCTT GCGACCTTCC
-------------------------------------------------------------
+2 A  A  G  H   Y  Q  A   E  A  A  L   L  Q  C   T  A  D
     ----------------------------------------------------
3801 CGGCGGGCCA TTACCAGGCC GAAGCAGCGT TGTTGCAGTG CACGGCAGAT
     GCCGCCCGGT AATGGTCCGG CTTCGTCGCA ACAACGTCAC GTGCCGTCTA
-------------------------------------------------------------
+2   T  L  A  D   A  V  L   I  T  T   A  H  A  W   Q  H  Q
     ----------------------------------------------------
3851 ACACTTGCTG ATGCGGTGCT GATTACGACC GCTCACGCGT GGCAGCATCA
     TGTGAACGAC TACGCCACGA CTAATGCTGG CGAGTGCGCA CCGTCGTAGT
-------------------------------------------------------------
+2   G  K  T   L  F  I  S   R  K  T   Y  R  I   D  G  S
     ----------------------------------------------------
3901 GGGGAAAACC TTATTTATCA GCCGGAAAAC CTACCGGATT GATGGTAGTG
     CCCCTTTTGG AATAAATAGT CGGCCTTTTG GATGGCCTAA CTACCATCAC
-------------------------------------------------------------
+2 G  Q  M  A   I  T  V   D  V  E  V   A  S  D   T  P  H
     ----------------------------------------------------
3951 GTCAAATGGC GATTACCGTT GATGTTGAAG TGGCGAGCGA TACACCGCAT
     CAGTTTACCG CTAATGGCAA CTACAACTTC ACCGCTCGCT ATGTGGCGTA
-------------------------------------------------------------
+2   P  A  R  I   G  L  N   C  Q  L   A  Q  V  A   E  R  V
     ----------------------------------------------------
4001 CCGGCGCGGA TTGGCCTGAA CTGCCAGCTG GCGCAGGTAG CAGAGCGGGT
     GGCCGCGCCT AACCGGACTT GACGGTCGAC CGCGTCCATC GTCTCGCCCA
-------------------------------------------------------------
+2   N  W  L   G  L  G  P   Q  E  N   Y  P  D   R  L  T
     ----------------------------------------------------
4051 AAACTGGCTC GGATTAGGGC CGCAAGAAAA CTATCCCGAC CGCCTTACTG
     TTTGACCGAG CCTAATCCCG GCGTTCTTTT GATAGGGCTG GCGGAATGAC
-------------------------------------------------------------
```

```
+2 A  A  C  F  D  R  W  D  L  P  L  S  D  M  Y  T  P
   ---------------------------------------------------
4101 CCGCCTGTTT TGACCGCTGG GATCTGCCAT TGTCAGACAT GTATACCCCG
     GGCGGACAAA ACTGGCGACC CTAGACGGTA ACAGTCTGTA CATATGGGGC
-------------------------------------------------------------
+2 Y  V  F  P  S  E  N  G  L  R  C  G  T  R  E  L  N
   ---------------------------------------------------
4151 TACGTCTTCC CGAGCGAAAA CGGTCTGCGC TGCGGGACGC GCGAATTGAA
     ATGCAGAAGG GCTCGCTTTT GCCAGACGCG ACGCCCTGCG CGCTTAACTT
-------------------------------------------------------------
+2 Y  G  P  H  Q  W  R  G  D  F  Q  F  N  I  S  R
   ---------------------------------------------------
4201 TTATGGCCCA CACCAGTGGC GCGGCGACTT CCAGTTCAAC ATCAGCCGCT
     AATACCGGGT GTGGTCACCG CGCCGCTGAA GGTCAAGTTG TAGTCGGCGA
-------------------------------------------------------------
+2 Y  S  Q  Q  Q  L  M  E  T  S  H  R  H  L  L  H  A
   ---------------------------------------------------
4251 ACAGTCAACA GCAACTGATG GAAACCAGCC ATCGCCATCT GCTGCACGCG
     TGTCAGTTGT CGTTGACTAC CTTTGGTCGG TAGCGGTAGA CGACGTGCGC
-------------------------------------------------------------
+2 E  E  G  T  W  L  N  I  D  G  F  H  M  G  I  G  G
   ---------------------------------------------------
4301 GAAGAAGGCA CATGGCTGAA TATCGACGGT TTCCATATGG GGATTGGTGG
     CTTCTTCCGT GTACCGACTT ATAGCTGCCA AAGGTATACC CCTAACCACC
-------------------------------------------------------------
+2 D  D  S  W  S  P  S  V  S  A  E  F  Q  L  S  A
   ---------------------------------------------------
4351 CGACGACTCC TGGAGCCCGT CAGTATCGGC GGAATTCCAG CTGAGCGCCG
     GCTGCTGAGG ACCTCGGGCA GTCATAGCCG CCTTAAGGTC GACTCGCGGC
-------------------------------------------------------------
+2 G  R  Y  H  Y  Q  L  V  W  C  Q  K  R  S  D  Y  K
   ---------------------------------------------------
4401 GTCGCTACCA TTACCAGTTG GTCTGGTGTC AAAAAAGATC TGACTATAAA
     CAGCGATGGT AATGGTCAAC CAGACCACAG TTTTTTCTAG ACTGATATTT
-------------------------------------------------------------
+2 D  E  D  L  D  H  H  H  H  H  H  R
   ------------------------------------->
4451 GATGAGGACC TCGACCATCA TCATCATCAT CACCGGTAAT AATAGGTAGA
     CTACTCCTGG AGCTGGTAGT AGTAGTAGTA GTGGCCATTA TTATCCATCT
-------------------------------------------------------------
4501 TAAGTGACTG ATTAGATGCA TTGATCCCTC GACCAATTCC GGTTATTTTC
     ATTCACTGAC TAATCTACGT AACTAGGGAG CTGGTTAAGG CCAATAAAAG
-------------------------------------------------------------
4551 CACCATATTG CCGTCTTTTG GCAATGTGAG GGCCCGGAAA CCTGGCCCTG
     GTGGTATAAC GGCAGAAAAC CGTTACACTC CCGGGCCTTT GGACCGGGAC
-------------------------------------------------------------
4601 TCTTCTTGAC GAGCATTCCT AGGGGTCTTT CCCCTCTCGC CAAAGGAATG
     AGAAGAACTG CTCGTAAGGA TCCCCAGAAA GGGGAGAGCG GTTTCCTTAC
-------------------------------------------------------------
4651 CAAGGTCTGT TGAATGTCGT GAAGGAAGCA GTTCCTCTGG AAGCTTCTTG
     GTTCCAGACA ACTTACAGCA CTTCCTTCGT CAAGGAGACC TTCGAAGAAC
-------------------------------------------------------------
4701 AAGACAAACA ACGTCTGTAG CGACCCTTTG CAGGCAGCGG AACCCCCCAC
     TTCTGTTTGT TGCAGACATC GCTGGGAAAC GTCCGTCGCC TTGGGGGGTG
-------------------------------------------------------------
4751 CTGGCGACAG GTGCCTCTGC GGCCAAAAGC CACGTGTATA AGATACACCT
     GACCGCTGTC CACGGAGACG CCGGTTTTCG GTGCACATAT TCTATGTGGA
-------------------------------------------------------------
```

```
4801  GCAAAGGCGG CACAACCCCA GTGCCACGTT GTGAGTTGGA TAGTTGTGGA
      CGTTTCCGCC GTGTTGGGGT CACGGTGCAA CACTCAACCT ATCAACACCT
---------------------------------------------------------------------
4851  AAGAGTCAAA TGGCTCTCCT CAAGCGTATT CAACAAGGGG CTGAAGGATG
      TTCTCAGTTT ACCGAGAGGA GTTCGCATAA GTTGTTCCCC GACTTCCTAC
---------------------------------------------------------------------
4901  CCCAGAAGGT ACCCCATTGT ATGGGATCTG ATCTGGGGCC TCGGTGCACA
      GGGTCTTCCA TGGGGTAACA TACCCTAGAC TAGACCCCGG AGCCACGTGT
---------------------------------------------------------------------
4951  TGCTTTACAT GTGTTTAGTC GAGGTTAAAA AACGTCTAGG CCCCCCGAAC
      ACGAAATGTA CACAAATCAG CTCCAATTTT TTGCAGATCC GGGGGGCTTG
---------------------------------------------------------------------
5001  CACGGGGACG TGCTTTTCCT TTGAAAAACA CGATGATAAT ACCATGATTG
      GTGCCCCTGC ACCAAAAGGA AACTTTTTGT GCTACTATTA TGGTACTAAC
---------------------------------------------------------------------
5051  AACAAGATGG ATTGCACGCA GGTTCTCCGG CCGCTTGGGT GGAGAGGCTA
      TTGTTCTACC TAACGTGCGT CCAAGAGGCC GGCGAACCCA CCTCTCCGAT
---------------------------------------------------------------------
5101  TTCGGCTATG ACTGGGCACA ACAGACAATC GGCTGCTCTG ATGCCGCCGT
      AAGCCGATAC TGACCCGTGT TGTCTGTTAG CCGACGAGAC TACGGCGGCA
---------------------------------------------------------------------
5151  GTTCCGGCTG TCAGCGCAGG GGCGCCCGGT TCTTTTTGTC AAGACCGACC
      CAAGGCCGAC AGTCGCGTCC CCGCGGGCCA AGAAAAACAG TTCTGGCTGG
---------------------------------------------------------------------
5201  TGTCCGGTGC CCTGAATGAA CTGCAGGACG AGGCAGCGCG GCTATCGTGG
      ACAGGCCACG GGACTTACTT GACGTCCTGC TCCGTCGCGC CGATAGCACC
---------------------------------------------------------------------
5251  CTGGCCACGA CGGGCGTTCC TTGCGCAGCT GTGCTCGACG TTGTCACTGA
      GACCGGTGCT GCCCGCAAGG AACGCGTCGA CACGAGCTGC AACAGTGACT
---------------------------------------------------------------------
5301  AGCGGGAAGG GACTGGCTGC TATTGGGCGA AGTGCCGGGG CAGGATCTCC
      TCGCCCTTCC CTGACCGACG ATAACCCGCT TCACGGCCCC GTCCTAGAGG
---------------------------------------------------------------------
5351  TGTCATCTCA CCTTGCTCCT GCCGAGAAAG TATCCATCAT GGCTGATGCA
      ACAGTAGAGT GGAACGAGGA CGGCTCTTTC ATAGGTAGTA CCGACTACGT
---------------------------------------------------------------------
5401  ATGCGGCGGC TGCATACGCT TGATCCGGCT ACCTGCCCAT TCGACCACCA
      TACGCCGCCG ACGTATGCGA ACTAGGCCGA TGGACGGGTA AGCTGGTGGT
---------------------------------------------------------------------
5451  AGCGAAACAT CGCATCGAGC GAGCACGTAC TCGGATGGAA GCCGGTCTTG
      TCGCTTTGTA GCGTAGCTCG CTCGTGCATG AGCCTACCTT CGGCCAGAAC
---------------------------------------------------------------------
5501  TCGATCAGGA TGATCTGGAC GAAGAGGATC AGGGGCTCGC GCCAGCCGAA
      AGCTAGTCCT ACTAGACCTG CTTCTCGTAG TCCCCGAGCG CGGTCGGCTT
---------------------------------------------------------------------
5551  CTGTTCGCCA GGCTCAAGGC GCGCATGCCC GACGGCGAGG ATCTCGTCGT
      GACAAGCGGT CCGAGTTCCG CGCGTACGGG CTGCCGCTCC TAGAGCAGCA
---------------------------------------------------------------------
5601  GACCCATGGC GATGCCTGCT TGCCGAATAT CATGGTGGAA AATGGCCGCT
      CTGGGTACCG CTACGGACGA ACGGCTTATA GTACCACCTT TTACCGGCGA
---------------------------------------------------------------------
5651  TTTCTGGATT CATCGACTGT GGCCGGCTGG GTGTGGCGGA CCGCTATCAG
      AAAGACCTAA GTAGCTGACA CCGGCCGACC CACACCGCCT GGCGATAGTC
---------------------------------------------------------------------
5701  GACATAGCGT TGGCTACCCG TGATATTGCT GAAGAGCTTG GCGGCGAATG
      CTGTATCGCA ACCGATGGGC ACTATAACGA CTTCTCGAAC CGCCGCTTAC
---------------------------------------------------------------------
```

```
5751  GGCTGACCGC TTCCTCGTGC TTTACGGTAT CGCCGCTCCC GATTCGCAGC
      CCGACTGGCG AAGGAGCACG AAATGCCATA GCGGCGAGGG CTAAGCGTCG
---------------------------------------------------------------------
5801  GCATCGCCTT CTATCGCCTT CTTGACGAGT TCTTCTGAGC GGGACTCTGG
      CGTAGCGGAA GATAGCGGAA GAACTGCTCA AGAAGACTCG CCCTGAGACC
---------------------------------------------------------------------
5851  GGTTCGCATC GATAAAATAA AAGATTTTAT TTAGTCTCCA GAAAAAGGGG
      CCAAGCGTAG CTATTTTATT TTCTAAAATA AATCAGAGGT CTTTTTCCCC
---------------------------------------------------------------------
5901  GGAATGAAAG ACCCCACCTG TAGGTTTGGC AAGCTAGCTT AAGTAACGCC
      CCTTACTTTC TGGGGTGGAC ATCCAAACCG TTCGATCGAA TTCATTGCGG
---------------------------------------------------------------------
5951  ATTTTGCAAG GCATGGAAAA ATACATAACT GAGAATAGAG AAGTTCAGAT
      TAAAACGTTC CGTACCTTTT TATGTATTGA CTCTTATCTC TTCAAGTCTA
---------------------------------------------------------------------
6001  CAAGGTCAGG AACAGATGGA ACAGCTGAAT ATGGGCCAAA CAGGATATCT
      GTTCCAGTCC TTGTCTACCT TGTCGACTTA TACCCGGTTT GTCCTATAGA
---------------------------------------------------------------------
6051  GTGGTAAGCA GTTCCTGCCC CGGCTCAGGG CCAAGAACAG ATGGAACAGC
      CACCATTCGT CAAGGACGGG GCCGAGTCCC GGTTCTTGTC TACCTTGTCG
---------------------------------------------------------------------
6101  TGAATATGGG CCAAACAGGA TATCTGTGGT AAGCAGTTCC TGCCCCGGCT
      ACTTATACCC GGTTTGTCCT ATAGACACCA TTCGTCAAGG ACGGGGCCGA
---------------------------------------------------------------------
6151  CAGGGCCAAG AACAGATGGT CCCCAGATGC GGTCCAGCCC TCAGCAGTTT
      GTCCCGGTTC TTGTCTACCA GGGGTCTACG CCAGGTCGGG AGTCGTCAAA
---------------------------------------------------------------------
6201  CTAGAGAACC ATCAGATGTT TCCAGGGTGC CCCAAGGACC TGAAATGACC
      GATCTCTTGG TAGTCTACAA AGGTCCCACG GGGTTCCTGG ACTTTACTGG
---------------------------------------------------------------------
6251  CTGTGCCTTA TTTGAACTAA CCAATCAGTT CGCTTCTCGC TTCTGTTCGC
      GACACGGAAT AAACTTGATT GGTTAGTCAA GCGAAGAGCG AAGACAAGCG
---------------------------------------------------------------------
6301  GCGCTTCTGC TCCCCGAGCT CAATAAAAGA GCCCACAACC CCTCACTCGG
      CGCGAAGACG AGGGGCTCGA GTTATTTTCT CGGGTGTTGG GGAGTGAGCC
---------------------------------------------------------------------
6351  GGCGCCAGTC CTCCGATTGA CTGAGTCGCC CGGGTACCCG TGTATCCAAT
      CCGCGGTCAG GAGGCTAACT GACTCAGCGG GCCCATGGGC ACATAGGTTA
---------------------------------------------------------------------
6401  AAACCCTCTT GCAGTTGCAT CCGACTTGTG GTCTCGCTGT TCCTTGGGAG
      TTTGGGAGAA CGTCAACGTA GGCTGAACAC CAGAGCGACA AGGAACCCTC
---------------------------------------------------------------------
6451  GGTCTCCTCT GAGTGATTGA CTACCCGTCA GCGGGGGTCT TTCATTCATG
      CCAGAGGAGA CTCACTAACT GATGGGCAGT CGCCCCCAGA AAGTAAGTAC
---------------------------------------------------------------------
6501  CAGCATGTAT CAAAATTAAT TTGGTTTTTT TTCTTAAGTA TTTACATTAA
      GTCGTACATA GTTTTAATTA AACCAAAAAA AAGAATTCAT AAATGTAATT
---------------------------------------------------------------------
6551  ATGGCCATAG TTGCATTAAT GAATCGGCCA ACGCGCGGGG AGAGGCGGTT
      TACCGGTATC AACGTAATTA CTTAGCCGGT TGCGCGCCCC TCTCCGCCAA
---------------------------------------------------------------------
6601  TGCGTATTGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
      ACGCATAACC GCGAGAAGGC GAAGGAGCGA GTGACTGAGC GACGCGAGCC
---------------------------------------------------------------------
6651  TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
      AGCAAGCCGA CGCCGCTCGC CATAGTCGAG TGAGTTTCCG CCATTATGCC
---------------------------------------------------------------------
```

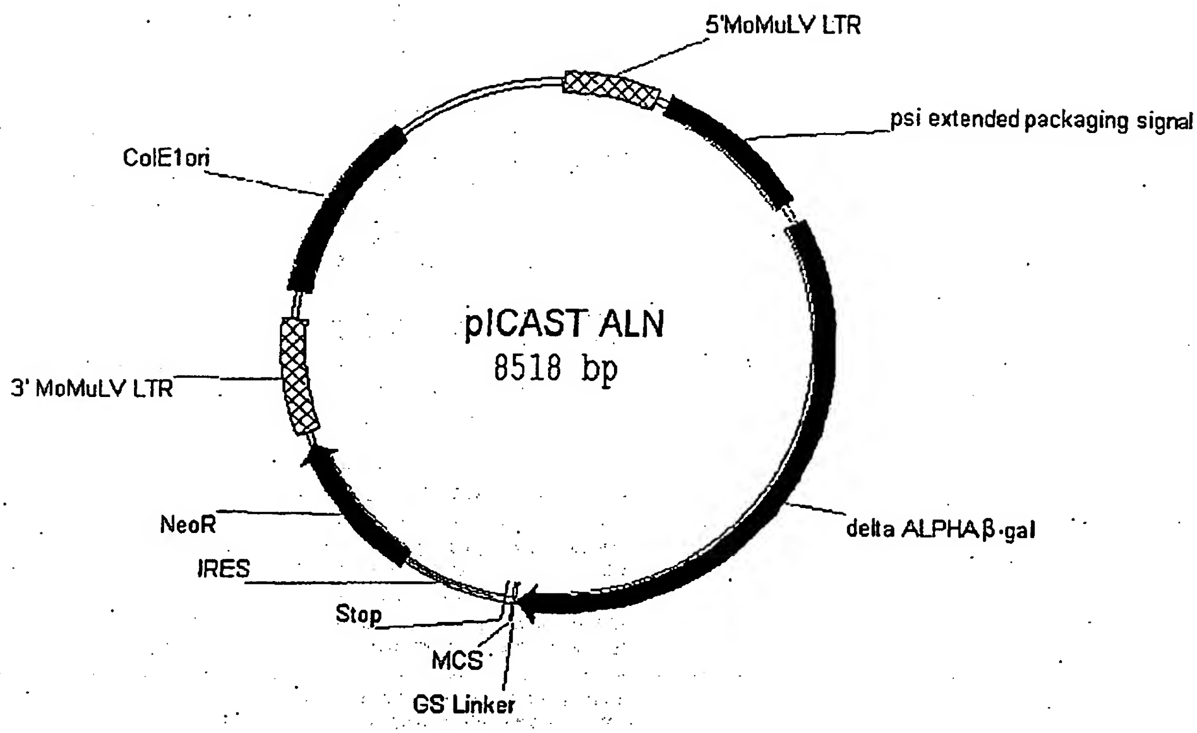

5'MoMuLV LTR
psi extended packaging signal
ColE1ori
pICAST ALN
8518 bp
3' MoMuLV LTR
NeoR
IRES
Stop
MCS
GS Linker
delta ALPHAβ-gal

Figure 11A

```
  1  CTGCAGCCTG AATATGGGCC AAACAGGATA TCTGTGGTAA GCAGTTCCTG
     GACGTCGGAC TTATACCCGG TTTGTCCTAT AGACACCATT CGTCAAGGAC

 51  CCCCGGCTCA GGGCCAAGAA CAGATGGAAC AGCTGAATAT GGGCCAAACA
     GGGGCCGAGT CCCGGTTCTT GTCTACCTTG TCGACTTATA CCCGGTTTGT

101  GGATATCTGT GGTAAGCAGT TCCTGCCCCG GCTCAGGGCC AAGAACAGAT
     CCTATAGACA CCATTCGTCA AGGACGGGGC CGAGTCCCGG TTCTTGTCTA

151  GGTCCCCAGA TGCGGTCCAG CCCTCAGCAG TTTCTAGAGA ACCATCAGAT
     CCAGGGGTCT ACGCCAGGTC GGGAGTCGTC AAAGATCTCT TGGTAGTCTA

201  GTTTCCAGGG TGCCCCAAGG ACCTGAAATG ACCCTGTGCC TTATTTGAAC
     CAAAGGTCCC ACGGGGTTCC TGGACTTTAC TGGGACACGG AATAAACTTG

251  TAACCAATCA GTTCGCTTCT CGCTTCTGTT CGCGCGCTTC TGCTCCCCGA
     ATTGGTTAGT CAAGCGAAGA GCGAAGACAA GCGCGCGAAG ACGAGGGGCT

301  GCTCAATAAA AGAGCCCACA ACCCCTCACT CGGGGCGCCA GTCCTCCGAT
     CGAGTTATTT TCTCGGGTGT TGGGGAGTGA GCCCCGCGGT CAGGAGGCTA

351  TGACTGAGTC GCCCGGGTAC CCGTGTATCC AATAAACCCT CTTGCAGTTG
     ACTGACTCAG CGGGCCCATG GGCACATAGG TTATTTGGGA GAACGTCAAC

401  CATCCGACTT GTGGTCTCGC TGTTCCTTGG GAGGGTCTCC TCTGAGTGAT
     GTAGGCTGAA CACCAGAGCG ACAAGGAACC CTCCCAGAGG AGACTCACTA

451  TGACTACCCG TCAGCGGGGG TCTTTCATTT GGGGGCTCGT CCGGGATCGG
     ACTGATGGGC AGTCGCCCCC AGAAAGTAAA CCCCCGAGCA GGCCCTAGCC

501  GAGACCCCTG CCCAGGGACC ACCGACCCAC CACCGGGAGG CAAGCTGGCC
     CTCTGGGGAC GGGTCCCTGG TGGCTGGGTG GTGGCCCTCC GTTCGACCGG

551  AGCAACTTAT CTGTGTCTGT CCGATTGTCT AGTGTCTATG ACTGATTTTA
     TCGTTGAATA GACACAGACA GGCTAACAGA TCACAGATAC TGACTAAAAT

601  TGCGCCTGCG TCGGTACTAG TTAGCTAACT AGCTCTGTAT CTGGCGGACC
     ACGCGGACGC AGCCATGATC AATCGATTGA TCGAGACATA GACCGCCTGG

651  CGTGGTGGAA CTGACGAGTT CTGAACACCC GGCCGCAACC CTGGGAGACG
     GCACCACCTT GACTGCTCAA GACTTGTGGG CCGGCGTTGG GACCCTCTGC

701  TCCCAGGGAC TTTGGGGGCC GTTTTTGTGG CCCGACCTGA GGAAGGGAGT
     AGGGTCCCTG AAACCCCCGG CAAAAACACC GGGCTGGACT CCTTCCCTCA

751  CGATGTGGAA TCCGACCCCG TCAGGATATG TGGTTCTGGT AGGAGACGAG
     GCTACACCTT AGGCTGGGGC AGTCCTATAC ACCAAGACCA TCCTCTGCTC

801  AACCTAAAAC AGTTCCCGCC TCCGTCTGAA TTTTTGCTTT CGGTTTGGAA
     TTGGATTTTG TCAAGGGCGG AGGCAGACTT AAAAACGAAA GCCAAACCTT

851  CCGAAGCCGC GCGTCTTGTC TGCTGCAGCA TCGTTCTGTG TTGTCTCTGT
     GGCTTCGGCG CGCAGAACAG ACGACGTCGT AGCAAGACAC AACAGAGACA

901  CTGACTGTGT TTCTGTATTT GTCTGAAAAT TAGGGCCAGA CTGTTACCAC
     GACTGACACA AAGACATAAA CAGACTTTTA ATCCCGGTCT GACAATGGTG
```

FIGURE 11B

```
951  TCCCTTAAGT TTGACCTTAG GTAACTGGAA AGATGTCGAG CGGCTCGCTC
     AGGGAATTCA AACTGGAATC CATTGACCTT TCTACAGCTC GCCGAGCGAG
---------------------------------------------------------------
1001 ACAACCAGTC GGTAGATGTC AAGAAGAGAC GTTGGGTTAC CTTCTGCTCT
     TGTTGGTCAG CCATCTACAG TTCTTCTCTG CAACCCAATG GAAGACGAGA
---------------------------------------------------------------
1051 GCAGAATGGC CAACCTTTAA CGTCGGATGG CCGCGAGACG GCACCTTTAA
     CGTCTTACCG GTTGGAAATT GCAGCCTACC GGCGCTCTGC CGTGGAAATT
---------------------------------------------------------------
1101 CCGAGACCTC ATCACCCAGG TTAAGATCAA GGTCTTTTCA CCTGGCCCGC
     GGCTCTGGAG TAGTGGGTCC AATTCTAGTT CCAGAAAAGT GGACCGGGCG
---------------------------------------------------------------
1151 ATGGACACCC AGACCAGGTC CCCTACATCG TGACCTGGGA AGCCTTGGCT
     TACCTGTGGG TCTGGTCCAG GGGATGTAGC ACTGGACCCT TCGGAACCGA
---------------------------------------------------------------
1201 TTTGACCCCC CTCCCTGGGT CAAGCCCTTT GTACACCCTA AGCCTCCGCC
     AAACTGGGGG GAGGGACCCA GTTCGGGAAA CATGTGGGAT TCGGAGGCGG
---------------------------------------------------------------
1251 TCCTCTTCCT CCATCCGCCC CGTCTCTCCC CCTTGAACCT CCTCGTTCGA
     AGGAGAAGGA GGTAGGCGGG GCAGAGAGGG GGAACTTGGA GGAGCAAGCT
---------------------------------------------------------------
1301 CCCCGCCTCG ATCCTCCCTT TATCCAGCCC TCACTCCTTC TCTAGGCGCC
     GGGGCGGAGC TAGGAGGGAA ATAGGTCGGG AGTGAGGAAG AGATCCGCGG
---------------------------------------------------------------
1351 GGCCGCTCTA GCCCATTAAT ACGACTCACT ATAGGGCGAT TCGAACACCA
     CCGGCGAGAT CGGGTAATTA TGCTGAGTGA TATCCCGCTA AGCTTGTGGT
---------------------------------------------------------------
1401 TGCACCATCA TCATCATCAC GTCGACTATA AAGATGAGGA CCTCGAGATG
     ACGTGGTAGT AGTAGTAGTG CAGCTGATAT TTCTACTCCT GGAGCTCTAC
---------------------------------------------------------------
1451 GGCGTGATTA CGGATTCACT GGCCGTCGTG GCCCGCACCG ATCGCCCTTC
     CCGCACTAAT GCCTAAGTGA CCGGCAGCAC CGGGCGTGGC TAGCGGGAAG
---------------------------------------------------------------
1501 CCAACAGTTA CGCAGCCTGA ATGGCGAATG GCGCTTTGCC TGGTTTCCGG
     GGTTGTCAAT GCGTCGGACT TACCGCTTAC CGCGAAACGG ACCAAAGGCC
---------------------------------------------------------------
1551 CACCAGAAGC GGTGCCGGAA AGCTGGCTGG AGTGCGATCT TCCTGAGGCC
     GTGGTCTTCG CCACGGCCTT TCGACCGACC TCACGCTAGA AGGACTCCGG
---------------------------------------------------------------
1601 GATACTGTCG TCGTCCCCTC AAACTGGCAG ATGCACGGTT ACGATGCGCC
     CTATGACAGC AGCAGGGGAG TTTGACCGTC TACGTGCCAA TGCTACGCGG
---------------------------------------------------------------
1651 CATCTACACC AACGTGACCT ATCCCATTAC GGTCAATCCG CCGTTTGTTC
     GTAGATGTGG TTGCACTGGA TAGGGTAATG CCAGTTAGGC GGCAAACAAG
---------------------------------------------------------------
1701 CCACGGAGAA TCCGACGGGT TGTTACTCGC TCACATTTAA TGTTGATGAA
     GGTGCCTCTT AGGCTGCCCA ACAATGAGCG AGTGTAAATT ACAACTACTT
---------------------------------------------------------------
1751 AGCTGGCTAC AGGAAGGCCA GACGCGAATT ATTTTTGATG GCGTTAACTC
     TCGACCGATG TCCTTCCGGT CTGCGCTTAA TAAAAACTAC CGCAATTGAG
---------------------------------------------------------------
1801 GGCGTTTCAT CTGTGGTGCA ACGGGCGCTG GGTCGGTTAC GGCCAGGACA
     CCGCAAAGTA GACACCACGT TGCCCGCGAC CCAGCCAATG CCGGTCCTGT
---------------------------------------------------------------
1851 GTCGTTTGCC GTCTGAATTT GACCTGAGCG CATTTTTACG CGCCGGAGAA
     CAGCAAACGG CAGACTTAAA CTGGACTCGC GTAAAAATGC GCGGCCTCTT
---------------------------------------------------------------
```

```
1901  AACCGCCTCG CGGTGATGGT GCTGCGCTGG AGTGACGGCA GTTATCTGGA
      TTGGCGGAGC GCCACTACCA CGACGCGACC TCACTGCCGT CAATAGACCT
-----------------------------------------------------------------------
1951  AGATCAGGAT ATGTGGCGGA TGAGCGGCAT TTTCCGTGAC GTCTCGTTGC
      TCTAGTCCTA TACACCGCCT ACTCGCCGTA AAAGGCACTG CAGAGCAACG
-----------------------------------------------------------------------
2001  TGCATAAACC GACTACACAA ATCAGCGATT TCCATGTTGC CACTCGCTTT
      ACGTATTTGG CTGATGTGTT TAGTCGCTAA AGGTACAACG GTGAGCGAAA
-----------------------------------------------------------------------
2051  AATGATGATT TCAGCCGCGC TGTACTGGAG GCTGAAGTTC AGATGTGCGG
      TTACTACTAA AGTCGGCGCG ACATGACCTC CGACTTCAAG TCTACACGCC
-----------------------------------------------------------------------
2101  CGAGTTGCGT GACTACCTAC GGGTAACAGT TTCTTTATGG CAGGGTGAAA
      GCTCAACGCA CTGATGGATG CCCATTGTCA AAGAAATACC GTCCCACTTT
-----------------------------------------------------------------------
2151  CGCAGGTCGC CAGCGGCACC GCGCCTTTCG GCGGTGAAAT TATCGATGAG
      GCGTCCAGCG GTCGCCGTGG CGCGGAAAGC CGCCACTTTA ATAGCTACTC
-----------------------------------------------------------------------
2201  CGTGGTGGTT ATGCCGATCG CGTCACACTA CGTCTGAACG TCGAAAACCC
      GCACCACCAA TACGGCTAGC GCAGTGTGAT GCAGACTTGC AGCTTTTGGG
-----------------------------------------------------------------------
2251  GAAACTGTGG AGCGCCGAAA TCCCGAATCT CTATCGTGCG GTGGTTGAAC
      CTTTGACACC TCGCGGCTTT AGGGCTTAGA GATAGCACGC CACCAACTTG
-----------------------------------------------------------------------
2301  TGCACACCGC CGACGGCACG CTGATTGAAG CAGAAGCCTG CGATGTCGGT
      ACGTGTGGCG GCTGCCGTGC GACTAACTTC GTCTTCGGAC GCTACAGCCA
-----------------------------------------------------------------------
2351  TTCCGCGAGG TGCGGATTGA AAATGGTCTG CTGCTGCTGA ACGGCAAGCC
      AAGGCGCTCC ACGCCTAACT TTTACCAGAC GACGACGACT TGCCGTTCGG
-----------------------------------------------------------------------
2401  GTTGCTGATT CGAGGCGTTA ACCGTCACGA GCATCATCCT CTGCATGGTC
      CAACGACTAA GCTCCGCAAT TGGCAGTGCT CGTAGTAGGA GACGTACCAG
-----------------------------------------------------------------------
2451  AGGTCATGGA TGAGCAGACG ATGGTGCAGG ATATCCTGCT GATGAAGCAG
      TCCAGTACCT ACTCGTCTGC TACCACGTCC TATAGGACGA CTACTTCGTC
-----------------------------------------------------------------------
2501  AACAACTTTA ACGCCGTGCG CTGTTCGCAT TATCCGAACC ATCCGCTGTG
      TTGTTGAAAT TGCGGCACGC GACAAGCGTA ATAGGCTTGG TAGGCGACAC
-----------------------------------------------------------------------
2551  GTACACGCTG TGCGACCGCT ACGGCCTGTA TGTGGTGGAT GAAGCCAATA
      CATGTGCGAC ACGCTGGCGA TGCCGGACAT ACACCACCTA CTTCGGTTAT
-----------------------------------------------------------------------
2601  TTGAAACCCA CGGCATGGTG CCAATGAATC GTCTGACCGA TGATCCGCGC
      AACTTTGGGT GCCGTACCAC GGTTACTTAG CAGACTGGCT ACTAGGCGCG
-----------------------------------------------------------------------
2651  TGGCTACCGG CGATGAGCGA ACGCGTAACG CGAATGGTGC AGCGCGATCG
      ACCGATGGCC GCTACTCGCT TGCGCATTGC GCTTACCACG TCGCGCTAGC
-----------------------------------------------------------------------
2701  TAATCACCCG AGTGTGATCA TCTGGTCGCT GGGGAATGAA TCAGGCCACG
      ATTAGTGGGC TCACACTAGT AGACCAGCGA CCCCTTACTT AGTCCGGTGC
-----------------------------------------------------------------------
2751  GCGCTAATCA CGACGCGCTG TATCGCTGGA TCAAATCTGT CGATCCTTCC
      CGCGATTAGT GCTGCGCGAC ATAGCGACCT AGTTTAGACA GCTAGGAAGG
-----------------------------------------------------------------------
2801  CGCCCGGTGC AGTATGAAGG CGGCGGAGCC GACACCACGG CCACCGATAT
      GCGGGCCACG TCATACTTCC GCCGCCTCGG CTGTGGTGCC GGTGGCTATA
-----------------------------------------------------------------------
```

```
2851  TATTTGCCCG ATGTACGCGC GCGTGGATGA AGACCAGCCC TTCCCGGCTG
      ATAAACGGGC TACATGCGCG CGCACCTACT TCTGGTCGGG AAGGGCCGAC
--------------------------------------------------------------
2901  TGCCGAAATG GTCCATCAAA AAATGGCTTT CGCTACCTGG AGAGACGCGC
      ACGGCTTTAC CAGGTAGTTT TTTACCGAAA GCGATGGACC TCTCTGCGCG
--------------------------------------------------------------
2951  CCGCTGATCC TTTGCGAATA CGCCCACGCG ATGGGTAACA GTCTTGGCGG
      GGCGACTAGG AAACGCTTAT GCGGGTGCGC TACCCATTGT CAGAACCGCC
--------------------------------------------------------------
3001  TTTCGCTAAA TACTGGCAGG CGTTTCGTCA GTATCCCCGT TTACAGGGCG
      AAAGCGATTT ATGACCGTCC GCAAAGCAGT CATAGGGGCA AATGTCCCGC
--------------------------------------------------------------
3051  GCTTCGTCTG GGACTGGGTG GATCAGTCGC TGATTAAATA TGATGAAAAC
      CGAAGCAGAC CCTGACCCAC CTAGTCAGCG ACTAATTTAT ACTACTTTTG
--------------------------------------------------------------
3101  GGCAACCCGT GGTCGGCTTA CGGCGGTGAT TTTGGCGATA CGCCGAACGA
      CCGTTGGGCA CCAGCCGAAT GCCGCCACTA AAACCGCTAT GCGGCTTGCT
--------------------------------------------------------------
3151  TCGCCAGTTC TGTATGAACG GTCTGGTCTT TGCCGACCGC ACGCCGCATC
      AGCGGTCAAG ACATACTTGC CAGACCAGAA ACGGCTGGCG TGCGGCGTAG
--------------------------------------------------------------
3201  CAGCGCTGAC GGAAGCAAAA CACCAGCAGC AGTTTTTCCA GTTCCGTTTA
      GTCGCGACTG CCTTCGTTTT GTGGTCGTCG TCAAAAAGGT CAAGGCAAAT
--------------------------------------------------------------
3251  TCCGGGCAAA CCATCGAAGT GACCAGCGAA TACCTGTTCC GTCATAGCGA
      AGGCCCGTTT GGTAGCTTCA CTGGTCGCTT ATGGACAAGG CAGTATCGCT
--------------------------------------------------------------
3301  TAACGAGCTC CTGCACTGGA TGGTGGCGCT GGATGGTAAG CCGCTGGCAA
      ATTGCTCGAG GACGTGACCT ACCACCGCGA CCTACCATTC GGCGACCGTT
--------------------------------------------------------------
3351  GCGGTGAAGT GCCTCTGGAT GTCGCTCCAC AAGGTAAACA GTTGATTGAA
      CGCCACTTCA CGGAGACCTA CAGCGAGGTG TTCCATTTGT CAACTAACTT
--------------------------------------------------------------
3401  CTGCCTGAAC TACCGCAGCC GGAGAGCGCC GGGCAACTCT GGCTCACAGT
      GACGGACTTG ATGGCGTCGG CCTCTCGCGG CCCGTTGAGA CCGAGTGTCA
--------------------------------------------------------------
3451  ACGCGTAGTG CAACCGAACG CGACCGCATG GTCAGAAGCC GGGCACATCA
      TGCGCATCAC GTTGGCTTGC GCTGGCGTAC CAGTCTTCGG CCCGTGTAGT
--------------------------------------------------------------
3501  GCGCCTGGCA GCAGTGGCGT CTGGCGGAAA ACCTCAGTGT GACGCTCCCC
      CGCGGACCGT CGTCACCGCA GACCGCCTTT TGGAGTCACA CTGCGAGGGG
--------------------------------------------------------------
3551  GCCGCGTCCC ACGCCATCCC GCATCTGACC ACCAGCGAAA TGGATTTTTG
      CGGCGCAGGG TGCGGTAGGG CGTAGACTGG TGGTCGCTTT ACCTAAAAAC
--------------------------------------------------------------
3601  CATCGAGCTG GGTAATAAGC GTTGGCAATT TAACCGCCAG TCAGGCTTTC
      GTAGCTCGAC CCATTATTCG CAACCGTTAA ATTGGCGGTC AGTCCGAAAG
--------------------------------------------------------------
3651  TTTCACAGAT GTGGATTGGC GATAAAAAAC AACTGCTGAC GCCGCTGCGC
      AAAGTGTCTA CACCTAACCG CTATTTTTTG TTGACGACTG CGGCGACGCG
--------------------------------------------------------------
3701  GATCAGTTCA CCCGTGCACC GCTGGATAAC GACATTGGCG TAAGTGAAGC
      CTAGTCAAGT GGGCACGTGG CGACCTATTG CTGTAACCGC ATTCACTTCG
--------------------------------------------------------------
3751  GACCCGCATT GACCCTAACG CCTGGGTCGA ACGCTGGAAG GCGGCGGGCC
      CTGGGCGTAA CTGGGATTGC GGACCCAGCT TGCGACCTTC CGCCGCCCGG
--------------------------------------------------------------
```

```
3801 ATTACCAGGC CGAAGCAGCG TTGTTGCAGT GCACGGCAGA TACACTTGCT
     TAATGGTCCG GCTTCGTCGC AACAACGTCA CGTGCCGTCT ATGTGAACGA
----------------------------------------------------------------------
3851 GATGCGGTGC TGATTACGAC CGCTCACGCG TGGCAGCATC AGGGGAAAAC
     CTACGCCACG ACTAATGCTG GCGAGTGCGC ACCGTCGTAG TCCCCTTTTG
----------------------------------------------------------------------
3901 CTTATTTATC AGCCGGAAAA CCTACCGGAT TGATGGTAGT GGTCAAATGG
     GAATAAATAG TCGGCCTTTT GGATGGCCTA ACTACCATCA CCAGTTTACC
----------------------------------------------------------------------
3951 CGATTACCGT TGATGTTGAA GTGGCGAGCG ATACACCGCA TCCGGCGCGG
     GCTAATGGCA ACTACAACTT CACCGCTCGC TATGTGGCGT AGGCCGCGCC
----------------------------------------------------------------------
4001 ATTGGCCTGA ACTGCCAGCT GGCGCAGGTA GCAGAGCGGG TAAACTGGCT
     TAACCGGACT TGACGGTCGA CCGCGTCCAT CGTCTCGCCC ATTTGACCGA
----------------------------------------------------------------------
4051 CGGATTAGGG CCGCAAGAAA ACTATCCCGA CCGCCTTACT GCCGCCTGTT
     GCCTAATCCC GGCGTTCTTT TGATAGGGCT GGCGGAATGA CGGCGGACAA
----------------------------------------------------------------------
4101 TTGACCGCTG GGATCTGCCA TTGTCAGACA TGTATACCCC GTACGTCTTC
     AACTGGCGAC CCTAGACGGT AACAGTCTGT ACATATGGGG CATGCAGAAG
----------------------------------------------------------------------
4151 CCGAGCGAAA ACGGTCTGCG CTGCGGGACG CGCGAATTGA ATTATGGCCC
     GGCTCGCTTT TGCCAGACGC GACGCCCTGC GCGCTTAACT TAATACCGGG
----------------------------------------------------------------------
4201 ACACCAGTGG CGCGGCGACT TCCAGTTCAA CATCAGCCGC TACAGTCAAC
     TGTGGTCACC GCGCCGCTGA AGGTCAAGTT GTAGTCGGCG ATGTCAGTTG
----------------------------------------------------------------------
4251 AGCAACTGAT GGAAACCAGC CATCGCCATC TGCTGCACGC GGAAGAAGGC
     TCGTTGACTA CCTTTGGTCG GTAGCGGTAG ACGACGTGCG CCTTCTTCCG
----------------------------------------------------------------------
4301 ACATGGCTGA ATATCGACGG TTTCCATATG GGGATTGGTG GCGACGACTC
     TGTACCGACT TATAGCTGCC AAAGGTATAC CCCTAACCAC CGCTGCTGAG
----------------------------------------------------------------------
4351 CTGGAGCCCG TCAGTATCGG CGGAATTCCA GCTGAGCGCC GGTCGCTACC
     GACCTCGGGC AGTCATAGCC GCCTTAAGGT CGACTCGCGG CCAGCGATGG
----------------------------------------------------------------------
4401 ATTACCAGTT GGTCTGGTGT CAAAAAAGAT CTGGAGGTGG TGGCAGCAGG
     TAATGGTCAA CCAGACCACA GTTTTTTCTA GACCTCCACC ACCGTCGTCC
----------------------------------------------------------------------
4451 CCTTGGCGCG CCGGATCCTT AATTAACAAT TGACCGGTAA TAATAGGTAG
     GGAACCGCGC GGCCTAGGAA TTAATTGTTA ACTGGCCATT ATTATCCATC
----------------------------------------------------------------------
4501 ATAAGTGACT GATTAGATGC ATTGATCCCT CGACCAATTC CGGTTATTTT
     TATTCACTGA CTAATCTACG TAACTAGGGA GCTGGTTAAG GCCAATAAAA
----------------------------------------------------------------------
4551 CCACCATATT GCCGTCTTTT GGCAATGTGA GGGCCCGGAA ACCTGGCCCT
     GGTGGTATAA CGGCAGAAAA CCGTTACACT CCCGGGCCTT TGGACCGGGA
----------------------------------------------------------------------
4601 GTCTTCTTGA CGAGCATTCC TAGGGGTCTT TCCCCTCTCG CCAAAGGAAT
     CAGAAGAACT GCTCGTAAGG ATCCCCAGAA AGGGGAGAGC GGTTTCCTTA
----------------------------------------------------------------------
4651 GCAAGGTCTG TTGAATGTCG TGAAGGAAGC AGTTCCTCTG GAAGCTTCTT
     CGTTCCAGAC AACTTACAGC ACTTCCTTCG TCAAGGAGAC CTTCGAAGAA
----------------------------------------------------------------------
4701 GAAGACAAAC AACGTCTGTA GCGACCCTTT GCAGGCAGCG GAACCCCCCA
     CTTCTGTTTG TTGCAGACAT CGCTGGGAAA CGTCCGTCGC CTTGGGGGGT
----------------------------------------------------------------------
```

```
4751  CCTGGCGACA GGTGCCTCTG CGGCCAAAAG CCACGTGTAT AAGATACACC
      GGACCGCTGT CCACGGAGAC GCCGGTTTTC GGTGCACATA TTCTATGTGG
-----------------------------------------------------------
4801  TGCAAAGGCG GCACAACCCC AGTGCCACGT TGTGAGTTGG ATAGTTGTGG
      ACGTTTCCGC CGTGTTGGGG TCACGGTGCA ACACTCAACC TATCAACACC
-----------------------------------------------------------
4851  AAAGAGTCAA ATGGCTCTCC TCAAGCGTAT TCAACAAGGG GCTGAAGGAT
      TTTCTCAGTT TACCGAGAGG AGTTCGCATA AGTTGTTCCC CGACTTCCTA
-----------------------------------------------------------
4901  GCCCAGAAGG TACCCCATTG TATGGGATCT GATCTGGGGC CTCGGTGCAC
      CGGGTCTTCC ATGGGGTAAC ATACCCTAGA CTAGACCCCG GAGCCACGTG
-----------------------------------------------------------
4951  ATGCTTTACA TGTGTTTAGT CGAGGTTAAA AAACGTCTAG GCCCCCCGAA
      TACGAAATGT ACACAAATCA GCTCCAATTT TTTGCAGATC CGGGGGGCTT
-----------------------------------------------------------
5001  CCACGGGGAC GTGGTTTTCC TTTGAAAAAC ACGATGATAA TACCATGATT
      GGTGCCCCTG CACCAAAAGG AAACTTTTTG TGCTACTATT ATGGTACTAA
-----------------------------------------------------------
5051  GAACAAGATG GATTGCACGC AGGTTCTCCG GCCGCTTGGG TGGAGAGGCT
      CTTGTTCTAC CTAACGTGCG TCCAAGAGGC CGGCGAACCC ACCTCTCCGA
-----------------------------------------------------------
5101  ATTCGGCTAT GACTGGGCAC AACAGACAAT CGGCTGCTCT GATGCCGCCG
      TAAGCCGATA CTGACCCGTG TTGTCTGTTA GCCGACGAGA CTACGGCGGC
-----------------------------------------------------------
5151  TGTTCCGGCT GTCAGCGCAG GGGCGCCCGG TTCTTTTTGT CAAGACCGAC
      ACAAGGCCGA CAGTCGCGTC CCCGCGGGCC AAGAAAAACA GTTCTGGCTG
-----------------------------------------------------------
5201  CTGTCCGGTG CCCTGAATGA ACTGCAGGAC GAGGCAGCGC GGCTATCGTG
      GACAGGCCAC GGGACTTACT TGACGTCCTG CTCCGTCGCG CCGATAGCAC
-----------------------------------------------------------
5251  GCTGGCCACG ACGGGCGTTC CTTGCGCAGC TGTGCTCGAC GTTGTCACTG
      CGACCGGTGC TGCCCGCAAG GAACGCGTCG ACACGAGCTG CAACAGTGAC
-----------------------------------------------------------
5301  AAGCGGGAAG GGACTGGCTG CTATTGGGCG AAGTGCCGGG GCAGGATCTC
      TTCGCCCTTC CCTGACCGAC GATAACCCGC TTCACGGCCC CGTCCTAGAG
-----------------------------------------------------------
5351  CTGTCATCTC ACCTTGCTCC TGCCGAGAAA GTATCCATCA TGGCTGATGC
      GACAGTAGAG TGGAACGAGG ACGGCTCTTT CATAGGTAGT ACCGACTACG
-----------------------------------------------------------
5401  AATGCGGCGG CTGCATACGC TTGATCCGGC TACCTGCCCA TTCGACCACC
      TTACGCCGCC GACGTATGCG AACTAGGCCG ATGGACGGGT AAGCTGGTGG
-----------------------------------------------------------
5451  AAGCGAAACA TCGCATCGAG CGAGCACGTA CTCGGATGGA AGCCGGTCTT
      TTCGCTTTGT AGCGTAGCTC GCTCGTGCAT GAGCCTACCT TCGGCCAGAA
-----------------------------------------------------------
5501  GTCGATCAGG ATGATCTGGA CGAAGAGCAT CAGGGGCTCG CGCCAGCCGA
      CAGCTAGTCC TACTAGACCT GCTTCTCGTA GTCCCCGAGC GCGGTCGGCT
-----------------------------------------------------------
5551  ACTGTTCGCC AGGCTCAAGG CGCGCATGCC CGACGGCGAG GATCTCGTCG
      TGACAAGCGG TCCGAGTTCC GCGCGTACGG GCTGCCGCTC CTAGAGCAGC
-----------------------------------------------------------
5601  TGACCCATGG CGATGCCTGC TTGCCGAATA TCATGGTGGA AAATGGCCGC
      ACTGGGTACC GCTACGGACG AACGGCTTAT AGTACCACCT TTTACCGGCG
-----------------------------------------------------------
5651  TTTTCTGGAT TCATCGACTG TGGCCGGCTG GGTGTGGCGG ACCGCTATCA
      AAAAGACCTA AGTAGCTGAC ACCGGCCGAC CCACACCGCC TGGCGATAGT
-----------------------------------------------------------
```

```
5701  GGACATAGCG TTGGCTACCC GTGATATTGC TGAAGAGCTT GGCGGCGAAT
      CCTGTATCGC AACCGATGGG CACTATAACG ACTTCTCGAA CCGCCGCTTA
--------------------------------------------------------------------
5751  GGGCTGACCG CTTCCTCGTG CTTTACGGTA TCGCCGCTCC CGATTCGCAG
      CCCGACTGGC GAAGGAGCAC GAAATGCCAT AGCGGCGAGG GCTAAGCGTC
--------------------------------------------------------------------
5801  CGCATCGCCT TCTATCGCCT TCTTGACGAG TTCTTCTGAG CGGGACTCTG
      GCGTAGCGGA AGATAGCGGA AGAACTGCTC AAGAAGACTC GCCCTGAGAC
--------------------------------------------------------------------
5851  GGGTTCGCAT CGATAAAATA AAAGATTTTA TTTAGTCTCC AGAAAAAGGG
      CCCAAGCGTA GCTATTTTAT TTTCTAAAAT AAATCAGAGG TCTTTTTCCC
--------------------------------------------------------------------
5901  GGGAATGAAA GACCCCACCT GTAGGTTTGG CAAGCTAGCT TAAGTAACGC
      CCCTTACTTT CTGGGGTGGA CATCCAAACC GTTCGATCGA ATTCATTGCG
--------------------------------------------------------------------
5951  CATTTTGCAA GGCATGGAAA AATACATAAC TGAGAATAGA GAAGTTCAGA
      GTAAAACGTT CCGTACCTTT TTATGTATTG ACTCTTATCT CTTCAAGTCT
--------------------------------------------------------------------
6001  TCAAGGTCAG GAACAGATGG AACAGCTGAA TATGGGCCAA ACAGGATATC
      AGTTCCAGTC CTTGTCTACC TTGTCGACTT ATACCCGGTT TGTCCTATAG
--------------------------------------------------------------------
6051  TGTGGTAAGC AGTTCCTGCC CCGGCTCAGG GCCAAGAACA GATGGAACAG
      ACACCATTCG TCAAGGACGG GGCCGAGTCC CGGTTCTTGT CTACCTTGTC
--------------------------------------------------------------------
6101  CTGAATATGG GCCAAACAGG ATATCTGTGG TAAGCAGTTC CTGCCCCGGC
      GACTTATACC CGGTTTGTCC TATAGACACC ATTCGTCAAG GACGGGGCCG
--------------------------------------------------------------------
6151  TCAGGGCCAA GAACAGATGG TCCCCAGATG CGGTCCAGCC CTCAGCAGTT
      AGTCCCGGTT CTTGTCTACC AGGGGTCTAC GCCAGGTCGG GAGTCGTCAA
--------------------------------------------------------------------
6201  TCTAGAGAAC CATCAGATGT TTCCAGGGTG CCCCAAGGAC CTGAAATGAC
      AGATCTCTTG GTAGTCTACA AAGGTCCCAC GGGGTTCCTG GACTTTACTG
--------------------------------------------------------------------
6251  CCTGTGCCTT ATTTGAACTA ACCAATCAGT TCGCTTCTCG CTTCTGTTCG
      GGACACGGAA TAAACTTGAT TGGTTAGTCA AGCGAAGAGC GAAGACAAGC
--------------------------------------------------------------------
6301  CGCGCTTCTG CTCCCCGAGC TCAATAAAAG AGCCCACAAC CCCTCACTCG
      GCGCGAAGAC GAGGGGCTCG AGTTATTTTC TCGGGTGTTG GGGAGTGAGC
--------------------------------------------------------------------
6351  GGGCGCCAGT CCTCCGATTG ACTGAGTCGC CCGGGTACCC GTGTATCCAA
      CCCGCGGTCA GGAGGCTAAC TGACTCAGCG GGCCCATGGG CACATAGGTT
--------------------------------------------------------------------
6401  TAAACCCTCT TGCAGTTGCA TCCGACTTGT GGTCTCGCTG TTCCTTGGGA
      ATTTGGGAGA ACGTCAACGT AGGCTGAACA CCAGAGCGAC AAGGAACCCT
--------------------------------------------------------------------
6451  GGGTCTCCTC TGAGTGATTG ACTACCCGTC AGCGGGGGTC TTTCATTCAT
      CCCAGAGGAG ACTCACTAAC TGATGGGCAG TCGCCCCCAG AAAGTAAGTA
--------------------------------------------------------------------
6501  GCAGCATGTA TCAAAATTAA TTTGGTTTTT TTTCTTAAGT ATTTACATTA
      CGTCGTACAT AGTTTTAATT AAACCAAAAA AAAGAATTCA TAAATGTAAT
--------------------------------------------------------------------
6551  AATGGCCATA GTTGCATTAA TGAATCGGCC AACGCGCGGG GAGAGGCGGT
      TTACCGGTAT CAACGTAATT ACTTAGCCGG TTGCGCGCCC CTCTCCGCCA
--------------------------------------------------------------------
6601  TTGCGTATTG GCGCTCTTCC GCTTCCTCGC TCACTGACTC GCTGCGCTCG
      AACGCATAAC CGCGAGAAGG CGAAGGAGCG AGTGACTGAG CGACGCGAGC
--------------------------------------------------------------------
```

```
6651  GTCGTTCGGC TGCGGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG
      CAGCAAGCCG ACGCCGCTCG CCATAGTCGA GTGAGTTTCC GCCATTATGC
-------------------------------------------------------------------------
6701  GTTATCCACA GAATCAGGGG ATAACGCAGG AAAGAACATG TGAGCAAAAG
      CAATAGGTGT CTTAGTCCCC TATTGCGTCC TTTCTTGTAC ACTCGTTTTC
-------------------------------------------------------------------------
6751  GCCAGCAAAA GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT GGCGTTTTTC
      CGGTCGTTTT CCGGTCCTTG GCATTTTTCC GGCGCAACGA CCGCAAAAAG
-------------------------------------------------------------------------
6801  CATAGGCTCC GCCCCCCTGA CGAGCATCAC AAAAATCGAC GCTCAAGTCA
      GTATCCGAGG CGGGGGGACT GCTCGTAGTG TTTTTAGCTG CGAGTTCAGT
-------------------------------------------------------------------------
6851  GAGGTGGCGA AACCCGACAG GACTATAAAG ATACCAGGCG TTTCCCCCTG
      CTCCACCGCT TTGGGCTGTC CTGATATTTC TATGGTCCGC AAAGGGGGAC
-------------------------------------------------------------------------
6901  GAAGCTCCCT CGTGCGCTCT CCTGTTCCGA CCCTGCCGCT TACCGGATAC
      CTTCGAGGGA GCACGCGAGA GGACAAGGCT GGGACGGCGA ATGGCCTATG
-------------------------------------------------------------------------
6951  CTGTCCGCCT TTCTCCCTTC GGGAAGCGTG GCGCTTTCTC ATAGCTCACG
      GACAGGCGGA AAGAGGGAAG CCCTTCGCAC CGCGAAAGAG TATCGAGTGC
-------------------------------------------------------------------------
7001  CTGTAGGTAT CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG CTGGGCTGTG
      GACATCCATA GAGTCAAGCC ACATCCAGCA AGCGAGGTTC GACCCGACAC
-------------------------------------------------------------------------
7051  TGCACGAACC CCCCGTTCAG CCCGACCGCT GCGCCTTATC CGGTAACTAT
      ACGTGCTTGG GGGGCAAGTC GGGCTGGCGA CGCGGAATAG GCCATTGATA
-------------------------------------------------------------------------
7101  CGTCTTGAGT CCAACCCGGT AAGACACGAC TTATCGCCAC TGGCAGCAGC
      GCAGAACTCA GGTTGGGCCA TTCTGTGCTG AATAGCGGTG ACCGTCGTCG
-------------------------------------------------------------------------
7151  CACTGGTAAC AGGATTAGCA GAGCGAGGTA TGTAGGCGGT GCTACAGAGT
      GTGACCATTG TCCTAATCGT CTCGCTCCAT ACATCCGCCA CGATGTCTCA
-------------------------------------------------------------------------
7201  TCTTGAAGTG GTGGCCTAAC TACGGCTACA CTAGAAGAAC AGTATTTGGT
      AGAACTTCAC CACCGGATTG ATGCCGATGT GATCTTCTTG TCATAAACCA
-------------------------------------------------------------------------
7251  ATCTGCGCTC TGCTGAAGCC AGTTACCTTC GGAAAAAGAG TTGGTAGCTC
      TAGACGCGAG ACGACTTCGG TCAATGGAAG CCTTTTTCTC AACCATCGAG
-------------------------------------------------------------------------
7301  TTGATCCGGC AAACAAACCA CCGCTGGTAG CGGTGGTTTT TTTGTTTGCA
      AACTAGGCCG TTTGTTTGGT GGCGACCATC GCCACCAAAA AAACAAACGT
-------------------------------------------------------------------------
7351  AGCAGCAGAT TACGCGCAGA AAAAAAGGAT CTCAAGAAGA TCCTTTGATC
      TCGTCGTCTA ATGCGCGTCT TTTTTTCCTA GAGTTCTTCT AGGAAACTAG
-------------------------------------------------------------------------
7401  TTTTCTACGG GGTCTGACGC TCAGTGGAAC GAAAACTCAC GTTAAGGGAT
      AAAAGATGCC CCAGACTGCG AGTCACCTTG CTTTTGAGTG CAATTCCCTA
-------------------------------------------------------------------------
7451  TTTGGTCATG AGATTATCAA AAAGGATCTT CACCTAGATC CTTTTGCGGC
      AAACCAGTAC TCTAATAGTT TTTCCTAGAA GTGGATCTAG GAAAACGCCG
-------------------------------------------------------------------------
7501  CGCAAATCAA TCTAAAGTAT ATATGAGTAA ACTTGGTCTG ACAGTTACCA
      GCGTTTAGTT AGATTTCATA TATACTCATT TGAACCAGAC TGTCAATGGT
-------------------------------------------------------------------------
7551  ATGCTTAATC AGTGAGGCAC CTATCTCAGC GATCTGTCTA TTTCGTTCAT
      TACGAATTAG TCACTCCGTG GATAGAGTCG CTAGACAGAT AAAGCAAGTA
-------------------------------------------------------------------------
```

```
7601  CCATAGTTGC CTGACTCCCC GTCGTGTAGA TAACTACGAT ACGGGAGGGC
      GGTATCAACG GACTGAGGGG CAGCACATCT ATTGATGCTA TGCCCTCCCG
---------------------------------------------------------------
7651  TTACCATCTG GCCCCAGTGC TGCAATGATA CCGCGAGACC CACGCTCACC
      AATGGTAGAC CGGGGTCACG ACGTTACTAT GGCGCTCTGG GTGCGAGTGG
---------------------------------------------------------------
7701  GGCTCCAGAT TTATCAGCAA TAAACCAGCC AGCCGGAAGG GCCGAGCGCA
      CCGAGGTCTA AATAGTCGTT ATTTGGTCGG TCGGCCTTCC CGGCTCGCGT
---------------------------------------------------------------
7751  GAAGTGGTCC TGCAACTTTA TCCGCCTCCA TCCAGTCTAT TAATTGTTGC
      CTTCACCAGG ACGTTGAAAT AGGCGGAGGT AGGTCAGATA ATTAACAACG
---------------------------------------------------------------
7801  CGGGAAGCTA GAGTAAGTAG TTCGCCAGTT AATAGTTTGC GCAACGTTGT
      GCCCTTCGAT CTCATTCATC AAGCGGTCAA TTATCAAACG CGTTGCAACA
---------------------------------------------------------------
7851  TGCCATTGCT ACAGGCATCG TGGTGTCACG CTCGTCGTTT GGTATGGCTT
      ACGGTAACGA TGTCCGTAGC ACCACAGTGC GAGCAGCAAA CCATACCGAA
---------------------------------------------------------------
7901  CATTCAGCTC CGGTTCCCAA CGATCAAGGC GAGTTACATG ATCCCCCATG
      GTAAGTCGAG GCCAAGGGTT GCTAGTTCCG CTCAATGTAC TAGGGGGTAC
---------------------------------------------------------------
7951  TTGTGCAAAA AAGCGGTTAG CTCCTTCGGT CCTCCGATCG TTGTCAGAAG
      AACACGTTTT TTCGCCAATC GAGGAAGCCA GGAGGCTAGC AACAGTCTTC
---------------------------------------------------------------
8001  TAAGTTGGCC GCAGTGTTAT CACTCATGGT TATGGCAGCA CTGCATAATT
      ATTCAACCGG CGTCACAATA GTGAGTACCA ATACCGTCGT GACGTATTAA
---------------------------------------------------------------
8051  CTCTTACTGT CATGCCATCC GTAAGATGCT TTTCTGTGAC TGGTGAGTAC
      GAGAATGACA GTACGGTAGG CATTCTACGA AAAGACACTG ACCACTCATG
---------------------------------------------------------------
8101  TCAACCAAGT CATTCTGAGA ATAGTGTATG CGGCGACCGA GTTGCTCTTG
      AGTTGGTTCA GTAAGACTCT TATCACATAC GCCGCTGGCT CAACGAGAAC
---------------------------------------------------------------
8151  CCCGGCGTCA ATACGGGATA ATACCGCGCC ACATAGCAGA ACTTTAAAAG
      GGGCCGCAGT TATGCCCTAT TATGGCGCGG TGTATCGTCT TGAAATTTTC
---------------------------------------------------------------
8201  TGCTCATCAT TGGAAAACGT TCTTCGGGGC GAAAACTCTC AAGGATCTTA
      ACGAGTAGTA ACCTTTTGCA AGAAGCCCCG CTTTTGAGAG TTCCTAGAAT
---------------------------------------------------------------
8251  CCGCTGTTGA GATCCAGTTC GATGTAACCC ACTCGTGCAC CCAACTGATC
      GGCGACAACT CTAGGTCAAG CTACATTGGG TGAGCACGTG GGTTGACTAG
---------------------------------------------------------------
8301  TTCAGCATCT TTTACTTTCA CCAGCGTTTC TGGGTGAGCA AAAACAGGAA
      AAGTCGTAGA AAATGAAAGT GGTCGCAAAG ACCCACTCGT TTTTGTCCTT
---------------------------------------------------------------
8351  GGCAAAATGC CGCAAAAAAG GGAATAAGGG CGACACGGAA ATGTTGAATA
      CCGTTTTACG GCGTTTTTTC CCTTATTCCC GCTGTGCCTT TACAACTTAT
---------------------------------------------------------------
8401  CTCATACTCT TCCTTTTTCA ATATTATTGA AGCATTTATC AGGGTTATTG
      GAGTATGAGA AGGAAAAAGT TATAATAACT TCGTAAATAG TCCCAATAAC
---------------------------------------------------------------
8451  TCTCATGAGC GGATACATAT TTGAATGTAT TTAGAAAAAT AAACAAATAG
      AGAGTACTCG CCTATGTATA AACTTACATA AATCTTTTTA TTTGTTTATC
---------------------------------------------------------------
8501  GGGTTCCGCG CACATTTC
      CCCAAGGCGC GTGTAAAG
---------------------------------------------------------------
```

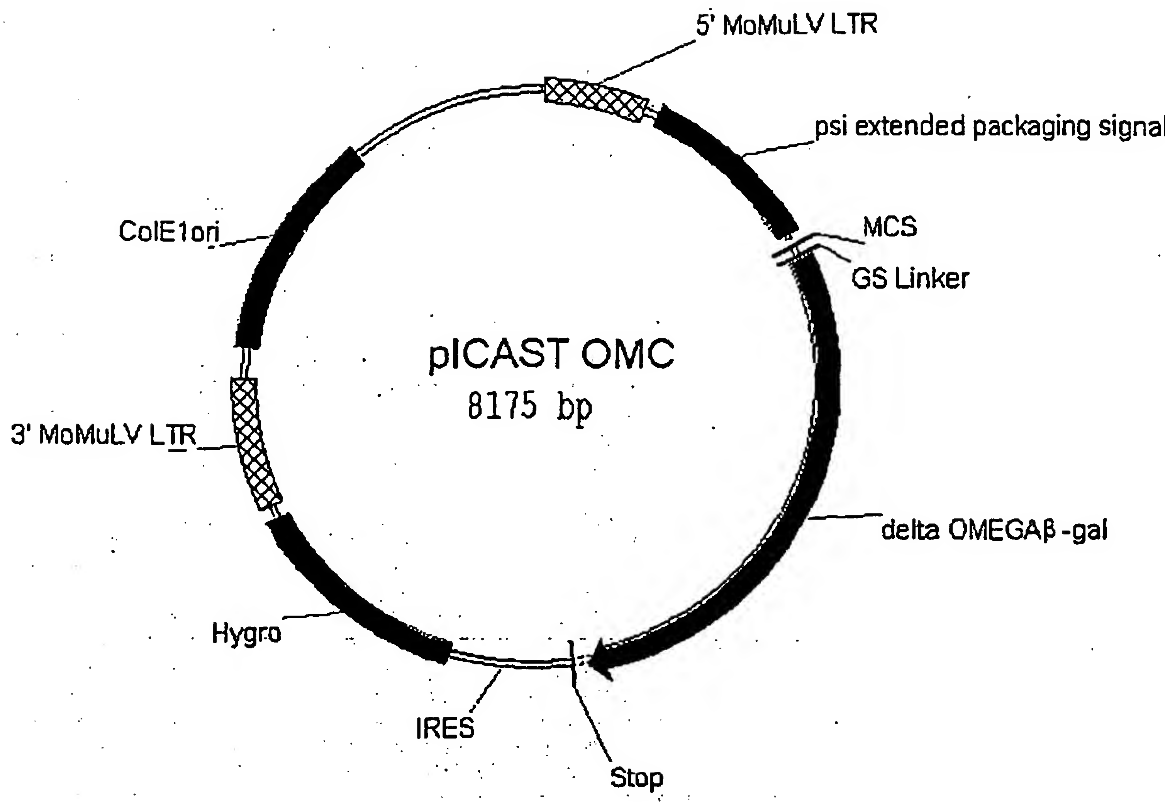

5' MoMuLV LTR
psi extended packaging signal
MCS
GS Linker
ColE1ori
pICAST OMC
8175 bp
3' MoMuLV LTR
delta OMEGAβ -gal
Hygro
IRES
Stop

Figure 12A

```
  1 CTGCAGCCTG AATATGGGCC AAACAGGATA TCTGTGGTAA GCAGTTCCTG
    GACGTCGGAC TTATACCCGG TTTGTCCTAT AGACACCATT CGTCAAGGAC
---------------------------------------------------------------
 51 CCCCGGCTCA GGGCCAAGAA CAGATGGAAC AGCTGAATAT GGGCCAAACA
    GGGGCCGAGT CCCGGTTCTT GTCTACCTTG TCGACTTATA CCCGGTTTGT
---------------------------------------------------------------
101 GGATATCTGT GGTAAGCAGT TCCTGCCCCG GCTCAGGGCC AAGAACAGAT
    CCTATAGACA CCATTCGTCA AGGACGGGGC CGAGTCCCGG TTCTTGTCTA
---------------------------------------------------------------
151 GGTCCCCAGA TGCGGTCCAG CCCTCAGCAG TTTCTAGAGA ACCATCAGAT
    CCAGGGGTCT ACGCCAGGTC GGGAGTCGTC AAAGATCTCT TGGTAGTCTA
---------------------------------------------------------------
201 GTTTCCAGGG TGCCCCAAGG ACCTGAAATG ACCCTGTGCC TTATTTGAAC
    CAAAGGTCCC ACGGGGTTCC TGGACTTTAC TGGGACACGG AATAAACTTG
---------------------------------------------------------------
251 TAACCAATCA GTTCGCTTCT CGCTTCTGTT CGCGCGCTTC TGCTCCCCGA
    ATTGGTTAGT CAAGCGAAGA GCGAAGACAA GCGCGCGAAG ACGAGGGGCT
---------------------------------------------------------------
301 GCTCAATAAA AGAGCCCACA ACCCCTCACT CGGGGCGCCA GTCCTCCGAT
    CGAGTTATTT TCTCGGGTGT TGGGGAGTGA GCCCCGCGGT CAGGAGGCTA
---------------------------------------------------------------
351 TGACTGAGTC GCCCGGGTAC CCGTGTATCC AATAAACCCT CTTGCAGTTG
    ACTGACTCAG CGGGCCCATG GGCACATAGG TTATTTGGGA GAACGTCAAC
---------------------------------------------------------------
401 CATCCGACTT GTGGTCTCGC TGTTCCTTGG GAGGGTCTCC TCTGAGTGAT
    GTAGGCTGAA CACCAGAGCG ACAAGGAACC CTCCCAGAGG AGACTCACTA
---------------------------------------------------------------
451 TGACTACCCG TCAGCGGGGG TCTTTCATTT GGGGGCTCGT CCGGGATCGG
    ACTGATGGGC AGTCGCCCCC AGAAAGTAAA CCCCCGAGCA GGCCCTAGCC
---------------------------------------------------------------
501 GAGACCCCTG CCCAGGGACC ACCGACCCAC CACCGGGAGG CAAGCTGGCC
    CTCTGGGGAC GGGTCCCTGG TGGCTGGGTG GTGGCCCTCC GTTCGACCGG
---------------------------------------------------------------
551 AGCAACTTAT CTGTGTCTGT CCGATTGTCT AGTGTCTATG ACTGATTTTA
    TCGTTGAATA GACACAGACA GGCTAACAGA TCACAGATAC TGACTAAAAT
---------------------------------------------------------------
601 TGCGCCTGCG TCGGTACTAG TTAGCTAACT AGCTCTGTAT CTGGCGGACC
    ACGCGGACGC AGCCATGATC AATCGATTGA TCGAGACATA GACCGCCTGG
---------------------------------------------------------------
651 CGTGGTGGAA CTGACGAGTT CTGAACACCC GGCCGCAACC CTGGGAGACG
    GCACCACCTT GACTGCTCAA GACTTGTGGG CCGGCGTTGG GACCCTCTGC
---------------------------------------------------------------
701 TCCCAGGGAC TTTGGGGGCC GTTTTTGTGG CCCGACCTGA GGAAGGGAGT
    AGGGTCCCTG AAACCCCCGG CAAAAACACC GGGCTGGACT CCTTCCCTCA
---------------------------------------------------------------
751 CGATGTGGAA TCCGACCCCG TCAGGATATG TGGTTCTGGT AGGAGACGAG
    GCTACACCTT AGGCTGGGGC AGTCCTATAC ACCAAGACCA TCCTCTGCTC
---------------------------------------------------------------
801 AACCTAAAAC AGTTCCCGCC TCCGTCTGAA TTTTTGCTTT CGGTTTGGAA
    TTGGATTTTG TCAAGGGCGG AGGCAGACTT AAAAACGAAA GCCAAACCTT
---------------------------------------------------------------
851 CCGAAGCCGC GCGTCTTGTC TGCTGCAGCA TCGTTCTGTG TTGTCTCTGT
    GGCTTCGGCG CGCAGAACAG ACGACGTCGT AGCAAGACAC AACAGAGACA
---------------------------------------------------------------
901 CTGACTGTGT TTCTGTATTT GTCTGAAAAT TAGGGCCAGA CTGTTACCAC
    GACTGACACA AAGACATAAA CAGACTTTTA ATCCCGGTCT GACAATGGTG
---------------------------------------------------------------
```

FIGURE 12B

```
 951  TCCCTTAAGT TTGACCTTAG GTAACTGGAA AGATGTCGAG CGGCTCGCTC
      AGGGAATTCA AACTGGAATC CATTGACCTT TCTACAGCTC GCCGAGCGAG
-----------------------------------------------------------------------
1001  ACAACCAGTC GGTAGATGTC AAGAAGAGAC GTTGGGTTAC CTTCTGCTCT
      TGTTGGTCAG CCATCTACAG TTCTTCTCTG CAACCCAATG GAAGACGAGA
-----------------------------------------------------------------------
1051  GCAGAATGGC CAACCTTTAA CGTCGGATGG CCGCGAGACG GCACCTTTAA
      CGTCTTACCG GTTGGAAATT GCAGCCTACC GGCGCTCTGC CGTGGAAATT
-----------------------------------------------------------------------
1101  CCGAGACCTC ATCACCCAGG TTAAGATCAA GGTCTTTTCA CCTGGCCCGC
      GGCTCTGGAG TAGTGGGTCC AATTCTAGTT CCAGAAAAGT GGACCGGGCG
-----------------------------------------------------------------------
1151  ATGGACACCC AGACCAGGTC CCCTACATCG TGACCTGGGA AGCCTTGGCT
      TACCTGTGGG TCTGGTCCAG GGGATGTAGC ACTGGACCCT TCGGAACCGA
-----------------------------------------------------------------------
1201  TTTGACCCCC CTCCCTGGGT CAAGCCCTTT GTACACCCTA AGCCTCCGCC
      AAACTGGGGG GAGGGACCCA GTTCGGGAAA CATGTGGGAT TCGGAGGCGG
-----------------------------------------------------------------------
1251  TCCTCTTCCT CCATCCGCCC CGTCTCTCCC CCTTGAACCT CCTCGTTCGA
      AGGAGAAGGA GGTAGGCGGG GCAGAGAGGG GGAACTTGGA GGAGCAAGCT
-----------------------------------------------------------------------
1301  CCCCGCCTCG ATCCTCCCTT TATCCAGCCC TCACTCCTTC TCTAGGCGCC
      GGGGCGGAGC TAGGAGGGAA ATAGGTCGGG AGTGAGGAAG AGATCCGCGG
-----------------------------------------------------------------------
1351  GGCCGCTCTA GCCCATTAAT ACGACTCACT ATAGGGCGAT TCGAATCAGG
      CCGGCGAGAT CGGGTAATTA TGCTGAGTGA TATCCCGCTA AGCTTAGTCC
-----------------------------------------------------------------------
1401  CCTTGGCGCG CCGGATCCTT AATTAAGCGC AATTGGGAGG TGGCGGTAGC
      GGAACCGCGC GGCCTAGGAA TTAATTCGCG TTAACCCTCC ACCGCCATCG
-----------------------------------------------------------------------
1451  CTCGAGATGG GCGTGATTAC GGATTCACTG GCCGTCGTTT TACAACGTCG
      GAGCTCTACC CGCACTAATG CCTAAGTGAC CGGCAGCAAA ATGTTGCAGC
-----------------------------------------------------------------------
1501  TGACTGGGAA AACCCTGGCG TTACCCAACT TAATCGCCTT GCAGCACATC
      ACTGACCCTT TTGGGACCGC AATGGGTTGA ATTAGCGGAA CGTCGTGTAG
-----------------------------------------------------------------------
1551  CCCCTTTCGC CAGCTGGCGT AATAGCGAAG AGGCCCGCAC CGATCGCCCT
      GGGGAAAGCG GTCGACCGCA TTATCGCTTC TCCGGGCGTG GCTAGCGGGA
-----------------------------------------------------------------------
1601  TCCCAACAGT TACGCAGCCT GAATGGCGAA TGGCGCTTTG CCTGGTTTCC
      AGGGTTGTCA ATGCGTCGGA CTTACCGCTT ACCGCGAAAC GGACCAAAGG
-----------------------------------------------------------------------
1651  GGCACCAGAA GCGGTGCCGG AAAGCTGGCT GGAGTGCGAT CTTCCTGAGG
      CCGTGGTCTT CGCCACGGCC TTTCGACCGA CCTCACGCTA GAAGGACTCC
-----------------------------------------------------------------------
1701  CCGATACTGT CGTCGTCCCC TCAAACTGGC AGATGCACGG TTACGATGCG
      GGCTATGACA GCAGCAGGGG AGTTTGACCG TCTACGTGCC AATGCTACGC
-----------------------------------------------------------------------
1751  CCCATCTACA CCAACGTGAC CTATCCCATT ACGGTCAATC CGCCGTTTGT
      GGGTAGATGT GGTTGCACTG GATAGGGTAA TGCCAGTTAG GCGGCAAACA
-----------------------------------------------------------------------
1801  TCCCACGGAG AATCCGACGG GTTGTTACTC GCTCACATTT AATGTTGATG
      AGGGTGCCTC TTAGGCTGCC CAACAATGAG CGAGTGTAAA TTACAACTAC
-----------------------------------------------------------------------
1851  AAAGCTGGCT ACAGGAAGGC CAGACGCGAA TTATTTTTGA TGGCGTTAAC
      TTTCGACCGA TGTCCTTCCG GTCTGCGCTT AATAAAAACT ACCGCAATTG
-----------------------------------------------------------------------
```

```
1901  TCGGCGTTTC ATCTGTGGTG CAACGGGCGC TGGGTCGGTT ACGGCCAGGA
      AGCCGCAAAG TAGACACCAC GTTGCCCGCG ACCCAGCCAA TGCCGGTCCT

1951  CAGTCGTTTG CCGTCTGAAT TTGACCTGAG CGCATTTTTA CGCGCCGGAG
      GTCAGCAAAC GGCAGACTTA AACTGGACTC GCGTAAAAAT GCGCGGCCTC

2001  AAAACCGCCT CGCGGTGATG GTGCTGCGCT GGAGTGACGG CAGTTATCTG
      TTTTGGCGGA GCGCCACTAC CACGACGCGA CCTCACTGCC GTCAATAGAC

2051  GAAGATCAGG ATATGTGGCG GATGAGCGGC ATTTTCCGTG ACGTCTCGTT
      CTTCTAGTCC TATACACCGC CTACTCGCCG TAAAAGGCAC TGCAGAGCAA

2101  GCTGCATAAA CCGACTACAC AAATCAGCGA TTTCCATGTT GCCACTCGCT
      CGACGTATTT GGCTGATGTG TTTAGTCGCT AAAGGTACAA CGGTGAGCGA

2151  TTAATGATGA TTTCAGCCGC GCTGTACTGG AGGCTGAAGT TCAGATGTGC
      AATTACTACT AAAGTCGGCG CGACATGACC TCCGACTTCA AGTCTACACG

2201  GGCGAGTTGC GTGACTACCT ACGGGTAACA GTTTCTTTAT GGCAGGGTGA
      CCGCTCAACG CACTGATGGA TGCCCATTGT CAAAGAAATA CCGTCCCACT

2251  AACGCAGGTC GCCAGCGGCA CCGCGCCTTT CGGCGGTGAA ATTATCGATG
      TTGCGTCCAG CGGTCGCCGT GGCGCGGAAA GCCGCCACTT TAATAGCTAC

2301  AGCGTGGTGG TTATGCCGAT CGCGTCACAC TACGTCTGAA CGTCGAAAAC
      TCGCACCACC AATACGGCTA GCGCAGTGTG ATGCAGACTT GCAGCTTTTG

2351  CCGAAACTGT GGAGCGCCGA AATCCCGAAT CTCTATCGTG CGGTGGTTGA
      GGCTTTGACA CCTCGCGGCT TTAGGGCTTA GAGATAGCAC GCCACCAACT

2401  ACTGCACACC GCCGACGGCA CGCTGATTGA AGCAGAAGCC TGCGATGTCG
      TGACGTGTGG CGGCTGCCGT GCGACTAACT TCGTCTTCGG ACGCTACAGC

2451  GTTTCCGCGA GGTGCGGATT GAAAATGGTC TGCTGCTGCT GAACGGCAAG
      CAAAGGCGCT CCACGCCTAA CTTTTACCAG ACGACGACGA CTTGCCGTTC

2501  CCGTTGCTGA TTCGAGGCGT TAACCGTCAC GAGCATCATC CTCTGCATGG
      GGCAACGACT AAGCTCCGCA ATTGGCAGTG CTCGTAGTAG GAGACGTACC

2551  TCAGGTCATG GATGAGCAGA CGATGGTGCA GGATATCCTG CTGATGAAGC
      AGTCCAGTAC CTACTCGTCT GCTACCACGT CCTATAGGAC GACTACTTCG

2601  AGAACAACTT TAACGCCGTG CGCTGTTCGC ATTATCCGAA CCATCCGCTG
      TCTTGTTGAA ATTGCGGCAC GCGACAAGCG TAATAGGCTT GGTAGGCGAC

2651  TGGTACACGC TGTGCGACCG CTACGGCCTG TATGTGGTGG ATGAAGCCAA
      ACCATGTGCG ACACGCTGGC GATGCCGGAC ATACACCACC TACTTCGGTT

2701  TATTGAAACC CACGGCATGG TGCCAATGAA TCGTCTGACC GATGATCCGC
      ATAACTTTGG GTGCCGTACC ACGGTTACTT AGCAGACTGG CTACTAGGCG

2751  GCTGGCTACC GGCGATGAGC GAACGCGTAA CGCGAATGGT GCAGCGCGAT
      CGACCGATGG CCGCTACTCG CTTGCGCATT GCGCTTACCA CGTCGCGCTA

2801  CGTAATCACC CGAGTGTGAT CATCTGGTCG CTGGGGAATG AATCAGGCCA
      GCATTAGTGG GCTCACACTA GTAGACCAGC GACCCCTTAC TTAGTCCGGT
```

2851 CGGCGCTAAT CACGACGCGC TGTATCGCTG GATCAAATCT GTCGATCCTT
GCCGCGATTA GTGCTGCGCG ACATAGCGAC CTAGTTTAGA CAGCTAGGAA

2901 CCCGCCCGGT GCAGTATGAA GGCGGCGGAG CCGACACCAC GGCCACCGAT
GGGCGGGCCA CGTCATACTT CCGCCGCCTC GGCTGTGGTG CCGGTGGCTA

2951 ATTATTTGCC CGATGTACGC GCGCGTGGAT GAAGACCAGC CCTTCCCGGC
TAATAAACGG GCTACATGCG CGCGCACCTA CTTCTGGTCG GGAAGGGCCG

3001 TGTGCCGAAA TGGTCCATCA AAAAATGGCT TTCGCTACCT GGAGAGACGC
ACACGGCTTT ACCAGGTAGT TTTTTACCGA AAGCGATGGA CCTCTCTGCG

3051 GCCCGCTGAT CCTTTGCGAA TACGCCCACG CGATGGGTAA CAGTCTTGGC
CGGGCGACTA GGAAACGCTT ATGCGGGTGC GCTACCCATT GTCAGAACCG

3101 GGTTTCGCTA AATACTGGCA GGCGTTTCGT CAGTATCCCC GTTTACAGGG
CCAAAGCGAT TTATGACCGT CCGCAAAGCA GTCATAGGGG CAAATGTCCC

3151 CGGCTTCGTC TGGGACTGGG TGGATCAGTC GCTGATTAAA TATGATGAAA
GCCGAAGCAG ACCCTGACCC ACCTAGTCAG CGACTAATTT ATACTACTTT

3201 ACGGCAACCC GTGGTCGGCT TACGGCGGTG ATTTTGGCGA TACGCCGAAC
TGCCGTTGGG CACCAGCCGA ATGCCGCCAC TAAAACCGCT ATGCGGCTTG

3251 GATCGCCAGT TCTGTATGAA CGGTCTGGTC TTTGCCGACC GCACGCCGCA
CTAGCGGTCA AGACATACTT GCCAGACCAG AAACGGCTGG CGTGCGGCGT

3301 TCCAGCGCTG ACGGAAGCAA AACACCAGCA GCAGTTTTTC CAGTTCCGTT
AGGTCGCGAC TGCCTTCGTT TTGTGGTCGT CGTCAAAAAG GTCAAGGCAA

3351 TATCCGGGCA AACCATCGAA GTGACCAGCG AATACCTGTT CCGTCATAGC
ATAGGCCCGT TTGGTAGCTT CACTGGTCGC TTATGGACAA GGCAGTATCG

3401 GATAACGAGC TCCTGCACTG GATGGTGGCG CTGGATGGTA AGCCGCTGGC
CTATTGCTCG AGGACGTGAC CTACCACCGC GACCTACCAT TCGGCGACCG

3451 AAGCGGTGAA GTGCCTCTGG ATGTCGCTCC ACAAGGTAAA CAGTTGATTG
TTCGCCACTT CACGGAGACC TACAGCGAGG TGTTCCATTT GTCAACTAAC

3501 AACTGCCTGA ACTACCGCAG CCGGAGAGCG CCGGGCAACT CTGGCTCACA
TTGACGGACT TGATGGCGTC GGCCTCTCGC GGCCCGTTGA GACCGAGTGT

3551 GTACGCGTAG TGCAACCGAA CGCGACCGCA TGGTCAGAAG CCGGGCACAT
CATGCGCATC ACGTTGGCTT GCGCTGGCGT ACCAGTCTTC GGCCCGTGTA

3601 CAGCGCCTGG CAGCAGTGGC GTCTGGCGGA AAACCTCAGT GTGACGCTCC
GTCGCGGACC GTCGTCACCG CAGACCGCCT TTTGGAGTCA CACTGCGAGG

3651 CCGCCGCGTC CCACGCCATC CCGCATCTGA CCACCAGCGA AATGGATTTT
GGCGGCGCAG GGTGCGGTAG GGCGTAGACT GGTGGTCGCT TTACCTAAAA

3701 TGCATCGAGC TGGGTAATAA GCGTTGGCAA TTTAACCGCC AGTCAGGCTT
ACGTAGCTCG ACCCATTATT CGCAACCGTT AAATTGGCGG TCAGTCCGAA

3751 TCTTTCACAG ATGTGGATTG GCGATAAAAA ACAACTGCTG ACGCCGCTGC
AGAAAGTGTC TACACCTAAC CGCTATTTTT TGTTGACGAC TGCGGCGACG

```
3801 GCGATCAGTT CACCCGTGTC GATAGATCTG AACAGAAACT CATTTCCGAA
     CGCTAGTCAA GTGGGCACAG CTATCTAGAC TTGTCTTTGA GTAAAGGCTT

3851 GAAGACCTAG TCGACCATCA TCATCATCAT CACCGGTAAT AATAGGTAGA
     CTTCTGGATC AGCTGGTAGT AGTAGTAGTA GTGGCCATTA TTATCCATCT

3901 TAAGTGACTG ATTAGATGCA TTTCGACTAG ATCCCTCGAC CAATTCCGGT
     ATTCACTGAC TAATCTACGT AAAGCTGATC TAGGGAGCTG GTTAAGGCCA

3951 TATTTTCCAC CATATTGCCG TCTTTTGGCA ATGTGAGGGC CCGGAAACCT
     ATAAAAGGTG GTATAACGGC AGAAAACCGT TACACTCCCG GGCCTTTGGA

4001 GGCCCTGTCT TCTTGACGAG CATTCCTAGG GGTCTTTCCC CTCTCGCCAA
     CCGGGACAGA AGAACTGCTC GTAAGGATCC CCAGAAAGGG GAGAGCGGTT

4051 AGGAATGCAA GGTCTGTTGA ATGTCGTGAA GGAAGCAGTT CCTCTGGAAG
     TCCTTACGTT CCAGACAACT TACAGCACTT CCTTCGTCAA GGAGACCTTC

4101 CTTCTTGAAG ACAAACAACG TCTGTAGCGA CCCTTTGCAG GCAGCGGAAC
     GAAGAACTTC TGTTTGTTGC AGACATCGCT GGGAAACGTC CGTCGCCTTG

4151 CCCCCACCTG GCGACAGGTG CCTCTGCGGC CAAAAGCCAC GTGTATAAGA
     GGGGGTGGAC CGCTGTCCAC GGAGACGCCG GTTTTCGGTG CACATATTCT

4201 TACACCTGCA AAGGCGGCAC AACCCCAGTG CCACGTTGTG AGTTGGATAG
     ATGTGGACGT TTCCGCCGTG TTGGGGTCAC GGTGCAACAC TCAACCTATC

4251 TTGTGGAAAG AGTCAAATGG CTCTCCTCAA GCGTATTCAA CAAGGGGCTG
     AACACCTTTC TCAGTTTACC GAGAGGAGTT CGCATAAGTT GTTCCCCGAC

4301 AAGGATGCCC AGAAGGTACC CCATTGTATG GGATCTGATC TGGGGCCTCG
     TTCCTACGGG TCTTCCATGG GGTAACATAC CCTAGACTAG ACCCCGGAGC

4351 GTGCACATGC TTTACATGTG TTTAGTCGAG GTTAAAAAAC GTCTAGGCCC
     CACGTGTACG AAATGTACAC AAATCAGCTC CAATTTTTTG CAGATCCGGG

4401 CCCGAACCAC GGGGACGTGG TTTTCCTTTG AAAAACACGA TGATAATACC
     GGGCTTGGTG CCCCTGCACC AAAAGGAAAC TTTTTGTGCT ACTATTATGG

4451 ATGAAAAAGC CTGAACTCAC CGCGACGTCT GTCGAGAAGT TTCTGATCGA
     TACTTTTTCG GACTTGAGTG GCGCTGCAGA CAGCTCTTCA AAGACTAGCT

4501 AAAGTTCGAC AGCGTCTCCG ACCTGATGCA GCTCTCGGAG GGCGAAGAAT
     TTTCAAGCTG TCGCAGAGGC TGGACTACGT CGAGAGCCTC CCGCTTCTTA

4551 CTCGTGCTTT CAGCTTCGAT GTAGGAGGGC GTGGATATGT CCTGCGGGTA
     GAGCACGAAA GTCGAAGCTA CATCCTCCCG CACCTATACA GGACGCCCAT

4601 AATAGCTGCG CCGATGGTTT CTACAAAGAT CGTTATGTTT ATCGGCACTT
     TTATCGACGC GGCTACCAAA GATGTTTCTA GCAATACAAA TAGCCGTGAA

4651 TGCATCGGCC GCGCTCCCGA TTCCGGAAGT GCTTGACATT GGGGAATTTA
     ACGTAGCCGG CGCGAGGGCT AAGGCCTTCA CGAACTGTAA CCCCTTAAAT

4701 GCGAGAGCCT GACCTATTGC ATCTCCCGCC GTGCACAGGG TGTCACGTTG
     CGCTCTCGGA CTGGATAACG TAGAGGGCGG CACGTGTCCC ACAGTGCAAC
```

```
4751  CAAGACCTGC CTGAAACCGA ACTGCCCGCT GTTCTGCAGC CGGTCGCGGA
      GTTCTGGACG GACTTTGGCT TGACGGGCGA CAAGACGTCG GCCAGCGCCT
-------------------------------------------------------------
4801  GGCCATGGAT GCGATCGCTG CGGCCGATCT TAGCCAGACG AGCGGGTTCG
      CCGGTACCTA CGCTAGCGAC GCCGGCTAGA ATCGGTCTGC TCGCCCAAGC
-------------------------------------------------------------
4851  GCCCATTCGG ACCGCAAGGA ATCGGTCAAT ACACTACATG GCGTGATTTC
      CGGGTAAGCC TGGCGTTCCT TAGCCAGTTA TGTGATGTAC CGCACTAAAG
-------------------------------------------------------------
4901  ATATGCGCGA TTGCTGATCC CCATGTGTAT CACTGGCAAA CTGTGATGGA
      TATACGCGCT AACGACTAGG GGTACACATA GTGACCGTTT GACACTACCT
-------------------------------------------------------------
4951  CGACACCGTC AGTGCGTCCG TCGCGCAGGC TCTCGATGAG CTGATGCTTT
      GCTGTGGCAG TCACGCAGGC AGCGCGTCCG AGAGCTACTC GACTACGAAA
-------------------------------------------------------------
5001  GGGCCGAGGA CTGCCCCGAA GTCCGGCACC TCGTGCACGC GGATTTCGGC
      CCCGGCTCCT GACGGGGCTT CAGGCCGTGG AGCACGTGCG CCTAAAGCCG
-------------------------------------------------------------
5051  TCCAACAATG TCCTGACGGA CAATGGCCGC ATAACAGCGG TCATTGACTG
      AGGTTGTTAC AGGACTGCCT GTTACCGGCG TATTGTCGCC AGTAACTGAC
-------------------------------------------------------------
5101  GAGCGAGGCG ATGTTCGGGG ATTCCCAATA CGAGGTCGCC AACATCTTCT
      CTCGCTCCGC TACAAGCCCC TAAGGGTTAT GCTCCAGCGG TTGTAGAAGA
-------------------------------------------------------------
5151  TCTGGAGGCC GTGGTTGGCT TGTATGGAGC AGCAGACGCG CTACTTCGAG
      AGACCTCCGG CACCAACCGA ACATACCTCG TCGTCTGCGC GATGAAGCTC
-------------------------------------------------------------
5201  CGGAGGCATC CGGAGCTTGC AGGATCGCCG CGGCTCCGGG CGTATATGCT
      GCCTCCGTAG GCCTCGAACG TCCTAGCGGC GCCGAGGCCC GCATATACGA
-------------------------------------------------------------
5251  CCGCATTGGT CTTGACCAAC TCTATCAGAG CTTGGTTGAC GGCAATTTCG
      GGCGTAACCA GAACTGGTTG AGATAGTCTC GAACCAACTG CCGTTAAAGC
-------------------------------------------------------------
5301  ATGATGCAGC TTGGGCGCAG GGTCGATGCG ACGCAATCGT CCGATCCGGA
      TACTACGTCG AACCCGCGTC CCAGCTACGC TGCGTTAGCA GGCTAGGCCT
-------------------------------------------------------------
5351  GCCGGGACTG TCGGGCGTAC ACAAATCGCC CGCAGAAGCG CGGCCGTCTG
      CGGCCCTGAC AGCCCGCATG TGTTTAGCGG GCGTCTTCGC GCCGGCAGAC
-------------------------------------------------------------
5401  GACCGATGGC TGTGTAGAAG TACTCGCCGA TAGTGGAAAC CGACGCCCCA
      CTGGCTACCG ACACATCTTC ATGAGCGGCT ATCACCTTTG GCTGCGGGGT
-------------------------------------------------------------
5451  GCACTCGTCC GAGGGCAAAG GAATAGAGTA GATGCCGACC GGGATCTATC
      CGTGAGCAGG CTCCCGTTTC CTTATCTCAT CTACGGCTGG CCCTAGATAG
-------------------------------------------------------------
5501  GATAAAATAA AAGATTTTAT TTAGTCTCCA GAAAAAGGGG GGAATGAAAG
      CTATTTTATT TTCTAAAATA AATCAGAGGT CTTTTTCCCC CCTTACTTTC
-------------------------------------------------------------
5551  ACCCCACCTG TAGGTTTGGC AAGCTAGCTT AAGTAACGCC ATTTTGCAAG
      TGGGGTGGAC ATCCAAACCG TTCGATCGAA TTCATTGCGG TAAAACGTTC
-------------------------------------------------------------
5601  GCATGGAAAA ATACATAACT GAGAATAGAG AAGTTCAGAT CAAGGTCAGG
      CGTACCTTTT TATGTATTGA CTCTTATCTC TTCAAGTCTA GTTCCAGTCC
-------------------------------------------------------------
5651  AACAGATGGA ACAGCTGAAT ATGGGCCAAA CAGGATATCT GTGGTAAGCA
      TTGTCTACCT TGTCGACTTA TACCCGGTTT GTCCTATAGA CACCATTCGT
-------------------------------------------------------------
```

```
5701 GTTCCTGCCC CGGCTCAGGG CCAAGAACAG ATGGAACAGC TGAATATGGG
     CAAGGACGGG GCCGAGTCCC GGTTCTTGTC TACCTTGTCG ACTTATACCC
-----------------------------------------------------------------
5751 CCAAACAGGA TATCTGTGGT AAGCAGTTCC TGCCCCGGCT CAGGGCCAAG
     GGTTTGTCCT ATAGACACCA TTCGTCAAGG ACGGGGCCGA GTCCCGGTTC
-----------------------------------------------------------------
5801 AACAGATGGT CCCCAGATGC GGTCCAGCCC TCAGCAGTTT CTAGAGAACC
     TTGTCTACCA GGGGTCTACG CCAGGTCGGG AGTCGTCAAA GATCTCTTGG
-----------------------------------------------------------------
5851 ATCAGATGTT TCCAGGGTGC CCCAAGGACC TGAAATGACC CTGTGCCTTA
     TAGTCTACAA AGGTCCCACG GGGTTCCTGG ACTTTACTGG GACACGGAAT
-----------------------------------------------------------------
5901 TTTGAACTAA CCAATCAGTT CGCTTCTCGC TTCTGTTCGC GCGCTTCTGC
     AAACTTGATT GGTTAGTCAA GCGAAGAGCG AAGACAAGCG CGCGAAGACG
-----------------------------------------------------------------
5951 TCCCCGAGCT CAATAAAAGA GCCCACAACC CCTCACTCGG GGCGCCAGTC
     AGGGGCTCGA GTTATTTTCT CGGGTGTTGG GGAGTGAGCC CCGCGGTCAG
-----------------------------------------------------------------
6001 CTCCGATTGA CTGAGTCGCC CGGGTACCCG TGTATCCAAT AAACCCTCTT
     GAGGCTAACT GACTCAGCGG GCCCATGGGC ACATAGGTTA TTTGGGAGAA
-----------------------------------------------------------------
6051 GCAGTTGCAT CCGACTTGTG GTCTCGCTGT TCCTTGGGAG GGTCTCCTCT
     CGTCAACGTA GGCTGAACAC CAGAGCGACA AGGAACCCTC CCAGAGGAGA
-----------------------------------------------------------------
6101 GAGTGATTGA CTACCCGTCA GCGGGGGTCT TTCATTCATG CAGCATGTAT
     CTCACTAACT GATGGGCAGT CGCCCCCAGA AAGTAAGTAC GTCGTACATA
-----------------------------------------------------------------
6151 CAAAATTAAT TTGGTTTTTT TTCTTAAGTA TTTACATTAA ATGGCCATAG
     GTTTTAATTA AACCAAAAAA AAGAATTCAT AAATGTAATT TACCGGTATC
-----------------------------------------------------------------
6201 TTGCATTAAT GAATCGGCCA ACGCGCGGGG AGAGGCGGTT TGCGTATTGG
     AACGTAATTA CTTAGCCGGT TGCGCGCCCC TCTCCGCCAA ACGCATAACC
-----------------------------------------------------------------
6251 CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG TCGTTCGGCT
     GCGAGAAGGC GAAGGAGCGA GTGACTGAGC GACGCGAGCC AGCAAGCCGA
-----------------------------------------------------------------
6301 GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG TTATCCACAG
     CGCCGCTCGC CATAGTCGAG TGAGTTTCCG CCATTATGCC AATAGGTGTC
-----------------------------------------------------------------
6351 AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG CCAGCAAAAG
     TTAGTCCCCT ATTGCGTCCT TTCTTGTACA CTCGTTTTCC GGTCGTTTTC
-----------------------------------------------------------------
6401 GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC ATAGGCTCCG
     CGGTCCTTGG CATTTTTCCG GCGCAACGAC CGCAAAAAGG TATCCGAGGC
-----------------------------------------------------------------
6451 CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG AGGTGGCGAA
     GGGGGGACTG CTCGTAGTGT TTTTAGCTGC GAGTTCAGTC TCCACCGCTT
-----------------------------------------------------------------
6501 ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG AAGCTCCCTC
     TGGGCTGTCC TGATATTTCT ATGGTCCGCA AAGGGGGACC TTCGAGGGAG
-----------------------------------------------------------------
6551 GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC TGTCCGCCTT
     CACGCGAGAG GACAAGGCTG GGACGGCGAA TGGCCTATGG ACAGGCGGAA
-----------------------------------------------------------------
6601 TCTCCCTTCG GGAAGCGTGG CGCTTTCTCA TAGCTCACGC TGTAGGTATC
     AGAGGGAAGC CCTTCGCACC GCGAAAGAGT ATCGAGTGCG ACATCCATAG
-----------------------------------------------------------------
```

```
6651  TCAGTTCGGT GTAGGTCGTT CGCTCCAAGC TGGGCTGTGT GCACGAACCC
      AGTCAAGCCA CATCCAGCAA GCGAGGTTCG ACCCGACACA CGTGCTTGGG
-------------------------------------------------------------
6701  CCCGTTCAGC CCGACCGCTG CGCCTTATCC GGTAACTATC GTCTTGAGTC
      GGGCAAGTCG GGCTGGCGAC GCGGAATAGG CCATTGATAG CAGAACTCAG
-------------------------------------------------------------
6751  CAACCCGGTA AGACACGACT TATCGCCACT GGCAGCAGCC ACTGGTAACA
      GTTGGGCCAT TCTGTGCTGA ATAGCGGTGA CCGTCGTCGG TGACCATTGT
-------------------------------------------------------------
6801  GGATTAGCAG AGCGAGGTAT GTAGGCGGTG CTACAGAGTT CTTGAAGTGG
      CCTAATCGTC TCGCTCCATA CATCCGCCAC GATGTCTCAA GAACTTCACC
-------------------------------------------------------------
6851  TGGCCTAACT ACGGCTACAC TAGAAGAACA GTATTTGGTA TCTGCGCTCT
      ACCGGATTGA TGCCGATGTG ATCTTCTTGT CATAAACCAT AGACGCGAGA
-------------------------------------------------------------
6901  GCTGAAGCCA GTTACCTTCG GAAAAAGAGT TGGTAGCTCT TGATCCGGCA
      CGACTTCGGT CAATGGAAGC CTTTTTCTCA ACCATCGAGA ACTAGGCCGT
-------------------------------------------------------------
6951  AACAAACCAC CGCTGGTAGC GGTGGTTTTT TTGTTTGCAA GCAGCAGATT
      TTGTTTGGTG GCGACCATCG CCACCAAAAA AACAAACGTT CGTCGTCTAA
-------------------------------------------------------------
7001  ACGCGCAGAA AAAAAGGATC TCAAGAAGAT CCTTTGATCT TTTCTACGGG
      TGCGCGTCTT TTTTTCCTAG AGTTCTTCTA GGAAACTAGA AAAGATGCCC
-------------------------------------------------------------
7051  GTCTGACGCT CAGTGGAACG AAAACTCACG TTAAGGGATT TTGGTCATGA
      CAGACTGCGA GTCACCTTGC TTTTGAGTGC AATTCCCTAA AACCAGTACT
-------------------------------------------------------------
7101  GATTATCAAA AAGGATCTTC ACCTAGATCC TTTTAAATTA AAAATGAAGT
      CTAATAGTTT TTCCTAGAAG TGGATCTAGG AAAATTTAAT TTTTACTTCA
-------------------------------------------------------------
7151  TTGCGGCCGC AAATCAATCT AAAGTATATA TGAGTAAACT TGGTCTGACA
      AACGCCGGCG TTTAGTTAGA TTTCATATAT ACTCATTTGA ACCAGACTGT
-------------------------------------------------------------
7201  GTTACCAATG CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT
      CAATGGTTAC GAATTAGTCA CTCCGTGGAT AGAGTCGCTA GACAGATAAA
-------------------------------------------------------------
7251  CGTTCATCCA TAGTTGCCTG ACTCCCCGTC GTGTAGATAA CTACGATACG
      GCAAGTAGGT ATCAACGGAC TGAGGGGCAG CACATCTATT GATGCTATGC
-------------------------------------------------------------
7301  GGAGGGCTTA CCATCTGGCC CCAGTGCTGC AATGATACCG CGAGACCCAC
      CCTCCCGAAT GGTAGACCGG GGTCACGACG TTACTATGGC GCTCTGGGTG
-------------------------------------------------------------
7351  GCTCACCGGC TCCAGATTTA TCAGCAATAA ACCAGCCAGC CGGAAGGGCC
      CGAGTGGCCG AGGTCTAAAT AGTCGTTATT TGGTCGGTCG GCCTTCCCGG
-------------------------------------------------------------
7401  GAGCGCAGAA GTGGTCCTGC AACTTTATCC GCCTCCATCC AGTCTATTAA
      CTCGCGTCTT CACCAGGACG TTGAAATAGG CGGAGGTAGG TCAGATAATT
-------------------------------------------------------------
7451  TTGTTGCCGG GAAGCTAGAG TAAGTAGTTC GCCAGTTAAT AGTTTGCGCA
      AACAACGGCC CTTCGATCTC ATTCATCAAG CGGTCAATTA TCAAACGCGT
-------------------------------------------------------------
7501  ACGTTGTTGC CATTGCTACA GGCATCGTGG TGTCACGCTC GTCGTTTGGT
      TGCAACAACG GTAACGATGT CCGTAGCACC ACAGTGCGAG CAGCAAACCA
-------------------------------------------------------------
7551  ATGGCTTCAT TCAGCTCCGG TTCCCAACGA TCAAGGCGAG TTACATGATC
      TACCGAAGTA AGTCGAGGCC AAGGGTTGCT AGTTCCGCTC AATGTACTAG
-------------------------------------------------------------
```

```
7601  CCCCATGTTG TGCAAAAAAG CGGTTAGCTC CTTCGGTCCT CCGATCGTTG
      GGGGTACAAC ACGTTTTTTC GCCAATCGAG GAAGCCAGGA GGCTAGCAAC
-------------------------------------------------------------------------------
7651  TCAGAAGTAA GTTGGCCGCA GTGTTATCAC TCATGGTTAT GGCAGCACTG
      AGTCTTCATT CAACCGGCGT CACAATAGTG AGTACCAATA CCGTCGTGAC
-------------------------------------------------------------------------------
7701  CATAATTCTC TTACTGTCAT GCCATCCGTA AGATGCTTTT CTGTGACTGG
      GTATTAAGAG AATGACAGTA CGGTAGGCAT TCTACGAAAA GACACTGACC
-------------------------------------------------------------------------------
7751  TGAGTACTCA ACCAAGTCAT TCTGAGAATA GTGTATGCGG CGACCGAGTT
      ACTCATGAGT TGGTTCAGTA AGACTCTTAT CACATACGCC GCTGGCTCAA
-------------------------------------------------------------------------------
7801  GCTCTTGCCC GGCGTCAATA CGGGATAATA CCGCGCCACA TAGCAGAACT
      CGAGAACGGG CCGCAGTTAT GCCCTATTAT GGCGCGGTGT ATCGTCTTGA
-------------------------------------------------------------------------------
7851  TTAAAAGTGC TCATCATTGG AAAACGTTCT TCGGGGCGAA AACTCTCAAG
      AATTTTCACG AGTAGTAACC TTTTGCAAGA AGCCCCGCTT TTGAGAGTTC
-------------------------------------------------------------------------------
7901  GATCTTACCG CTGTTGAGAT CCAGTTCGAT GTAACCCACT CGTGCACCCA
      CTAGAATGGC GACAACTCTA GGTCAAGCTA CATTGGGTGA GCACGTGGGT
-------------------------------------------------------------------------------
7951  ACTGATCTTC AGCATCTTTT ACTTTCACCA GCGTTTCTGG GTGAGCAAAA
      TGACTAGAAG TCGTAGAAAA TGAAAGTGGT CGCAAAGACC CACTCGTTTT
-------------------------------------------------------------------------------
8001  ACAGGAAGGC AAAATGCCGC AAAAAAGGGA ATAAGGGCGA CACGGAAATG
      TGTCCTTCCG TTTTACGGCG TTTTTTCCCT TATTCCCGCT GTGCCTTTAC
-------------------------------------------------------------------------------
8051  TTGAATACTC ATACTCTTCC TTTTTCAATA TTATTGAAGC ATTTATCAGG
      AACTTATGAG TATGAGAAGG AAAAAGTTAT AATAACTTCG TAAATAGTCC
-------------------------------------------------------------------------------
8101  GTTATTGTCT CATGAGCGGA TACATATTTG AATGTATTTA GAAAAATAAA
      CAATAACAGA GTACTCGCCT ATGTATAAAC TTACATAAAT CTTTTTATTT
-------------------------------------------------------------------------------
8151  CAAATAGGGG TTCCGCGCAC ATTTC
      GTTTATCCCC AAGGCGCGTG TAAAG
-------------------------------------------------------------------------------
```

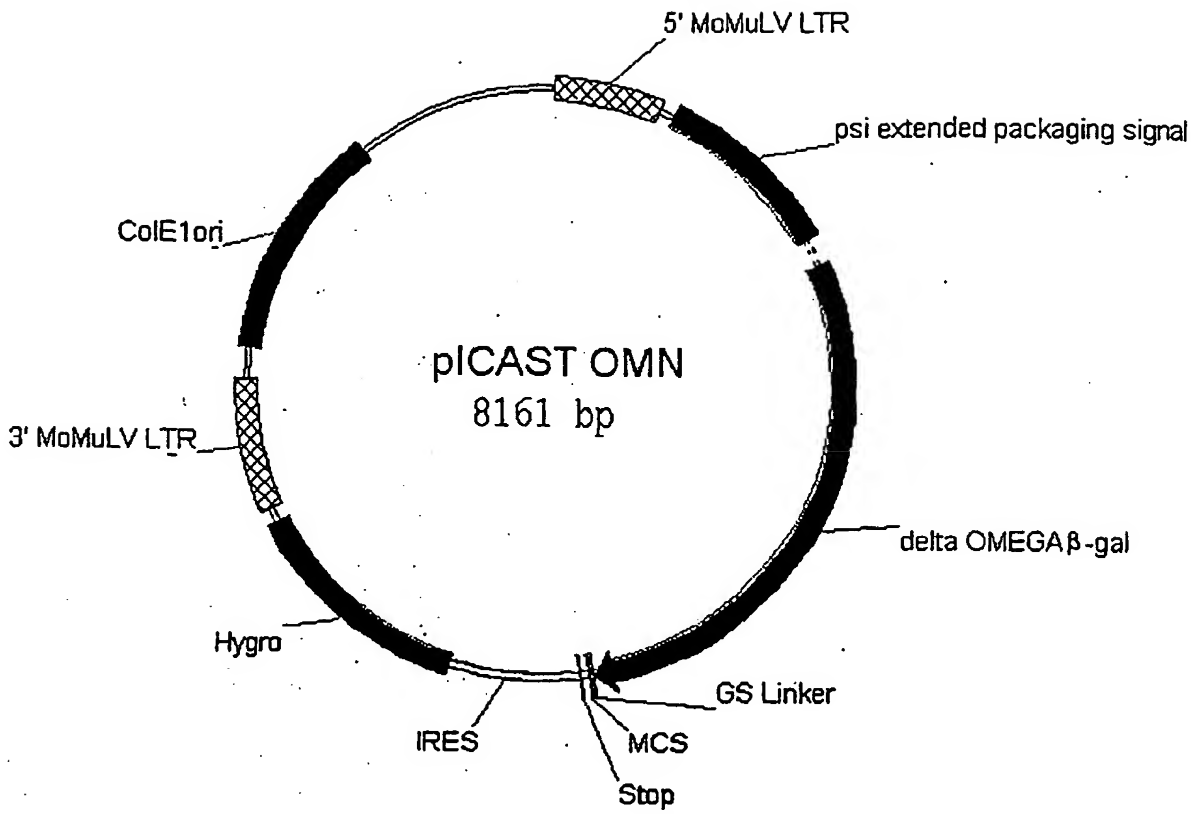

5' MoMuLV LTR
psi extended packaging signal
ColE1ori
pICAST OMN
8161 bp
3' MoMuLV LTR
delta OMEGAβ-gal
Hygro
GS Linker
IRES
MCS
Stop

Figure 13A

```
1   CTGCAGCCTG AATATGGGCC AAACAGGATA TCTGTGGTAA GCAGTTCCTG
    GACGTCGGAC TTATACCCGG TTTGTCCTAT AGACACCATT CGTCAAGGAC
---------------------------------------------------------------
51  CCCCGGCTCA GGGCCAAGAA CAGATGGAAC AGCTGAATAT GGGCCAAACA
    GGGGCCGAGT CCCGGTTCTT GTCTACCTTG TCGACTTATA CCCGGTTTGT
---------------------------------------------------------------
101 GGATATCTGT GGTAAGCAGT TCCTGCCCCG GCTCAGGGCC AAGAACAGAT
    CCTATAGACA CCATTCGTCA AGGACGGGGC CGAGTCCCGG TTCTTGTCTA
---------------------------------------------------------------
151 GGTCCCCAGA TGCGGTCCAG CCCTCAGCAG TTTCTAGAGA ACCATCAGAT
    CCAGGGGTCT ACGCCAGGTC GGGAGTCGTC AAAGATCTCT TGGTAGTCTA
---------------------------------------------------------------
201 GTTTCCAGGG TGCCCCAAGG ACCTGAAATG ACCCTGTGCC TTATTTGAAC
    CAAAGGTCCC ACGGGGTTCC TGGACTTTAC TGGGACACGG AATAAACTTG
---------------------------------------------------------------
251 TAACCAATCA GTTCGCTTCT CGCTTCTGTT CGCGCGCTTC TGCTCCCCGA
    ATTGGTTAGT CAAGCGAAGA GCGAAGACAA GCGCGCGAAG ACGAGGGGCT
---------------------------------------------------------------
301 GCTCAATAAA AGAGCCCACA ACCCCTCACT CGGGGCGCCA GTCCTCCGAT
    CGAGTTATTT TCTCGGGTGT TGGGGAGTGA GCCCCGCGGT CAGGAGGCTA
---------------------------------------------------------------
351 TGACTGAGTC GCCCGGGTAC CCGTGTATCC AATAAACCCT CTTGCAGTTG
    ACTGACTCAG CGGGCCCATG GGCACATAGG TTATTTGGGA GAACGTCAAC
---------------------------------------------------------------
401 CATCCGACTT GTGGTCTCGC TGTTCCTTGG GAGGGTCTCC TCTGAGTGAT
    GTAGGCTGAA CACCAGAGCG ACAAGGAACC CTCCCAGAGG AGACTCACTA
---------------------------------------------------------------
451 TGACTACCCG TCAGCGGGGG TCTTTCATTT GGGGGCTCGT CCGGGATCGG
    ACTGATGGGC AGTCGCCCCC AGAAAGTAAA CCCCCGAGCA GGCCCTAGCC
---------------------------------------------------------------
501 GAGACCCCTG CCCAGGGACC ACCGACCCAC CACCGGGAGG CAAGCTGGCC
    CTCTGGGGAC GGGTCCCTGG TGGCTGGGTG GTGGCCCTCC GTTCGACCGG
---------------------------------------------------------------
551 AGCAACTTAT CTGTGTCTGT CCGATTGTCT AGTGTCTATG ACTGATTTTA
    TCGTTGAATA GACACAGACA GGCTAACAGA TCACAGATAC TGACTAAAAT
---------------------------------------------------------------
601 TGCGCCTGCG TCGGTACTAG TTAGCTAACT AGCTCTGTAT CTGGCGGACC
    ACGCGGACGC AGCCATGATC AATCGATTGA TCGAGACATA GACCGCCTGG
---------------------------------------------------------------
651 CGTGGTGGAA CTGACGAGTT CTGAACACCC GGCCGCAACC CTGGGAGACG
    GCACCACCTT GACTGCTCAA GACTTGTGGG CCGGCGTTGG GACCCTCTGC
---------------------------------------------------------------
701 TCCCAGGGAC TTTGGGGGCC GTTTTTGTGG CCCGACCTGA GGAAGGGAGT
    AGGGTCCCTG AAACCCCCGG CAAAAACACC GGGCTGGACT CCTTCCCTCA
---------------------------------------------------------------
751 CGATGTGGAA TCCGACCCCG TCAGGATATG TGGTTCTGGT AGGAGACGAG
    GCTACACCTT AGGCTGGGGC AGTCCTATAC ACCAAGACCA TCCTCTGCTC
---------------------------------------------------------------
801 AACCTAAAAC AGTTCCCGCC TCCGTCTGAA TTTTTGCTTT CGGTTTGGAA
    TTGGATTTTG TCAAGGGCGG AGGCAGACTT AAAAACGAAA GCCAAACCTT
---------------------------------------------------------------
851 CCGAAGCCGC GCGTCTTGTC TGCTGCAGCA TCGTTCTGTG TTGTCTCTGT
    GGCTTCGGCG CGCAGAACAG ACGACGTCGT AGCAAGACAC AACAGAGACA
---------------------------------------------------------------
901 CTGACTGTGT TTCTGTATTT GTCTGAAAAT TAGGGCCAGA CTGTTACCAC
    GACTGACACA AAGACATAAA CAGACTTTTA ATCCCGGTCT GACAATGGTG
---------------------------------------------------------------
```

FIGURE 13B

```
951  TCCCTTAAGT TTGACCTTAG GTAACTGGAA AGATGTCGAG CGGCTCGCTC
     AGGGAATTCA AACTGGAATC CATTGACCTT TCTACAGCTC GCCGAGCGAG
---------------------------------------------------------------------
1001 ACAACCAGTC GGTAGATGTC AAGAAGAGAC GTTGGGTTAC CTTCTGCTCT
     TGTTGGTCAG CCATCTACAG TTCTTCTCTG CAACCCAATG GAAGACGAGA
---------------------------------------------------------------------
1051 GCAGAATGGC CAACCTTTAA CGTCGGATGG CCGCGAGACG GCACCTTTAA
     CGTCTTACCG GTTGGAAATT GCAGCCTACC GGCGCTCTGC CGTGGAAATT
---------------------------------------------------------------------
1101 CCGAGACCTC ATCACCCAGG TTAAGATCAA GGTCTTTTCA CCTGGCCCGC
     GGCTCTGGAG TAGTGGGTCC AATTCTAGTT CCAGAAAAGT GGACCGGGCG
---------------------------------------------------------------------
1151 ATGGACACCC AGACCAGGTC CCCTACATCG TGACCTGGGA AGCCTTGGCT
     TACCTGTGGG TCTGGTCCAG GGGATGTAGC ACTGGACCCT TCGGAACCGA
---------------------------------------------------------------------
1201 TTTGACCCCC CTCCCTGGGT CAAGCCCTTT GTACACCCTA AGCCTCCGCC
     AAACTGGGGG GAGGGACCCA GTTCGGGAAA CATGTGGGAT TCGGAGGCGG
---------------------------------------------------------------------
1251 TCCTCTTCCT CCATCCGCCC CGTCTCTCCC CCTTGAACCT CCTCGTTCGA
     AGGAGAAGGA GGTAGGCGGG GCAGAGAGGG GGAACTTGGA GGAGCAAGCT
---------------------------------------------------------------------
1301 CCCCGCCTCG ATCCTCCCTT TATCCAGCCC TCACTCCTTC TCTAGGCGCC
     GGGGCGGAGC TAGGAGGGAA ATAGGTCGGG AGTGAGGAAG AGATCCGCGG
---------------------------------------------------------------------
1351 GGCCGCTCTA GCCCATTAAT ACGACTCACT ATAGGGCGAT TCGAACACCA
     CCGGCGAGAT CGGGTAATTA TGCTGAGTGA TATCCCGCTA AGCTTGTGGT
---------------------------------------------------------------------
1401 TGCACCATCA TCATCATCAC GTCGACGAAC AGAAACTCAT TTCCGAAGAA
     ACGTGGTAGT AGTAGTAGTG CAGCTGCTTG TCTTTGAGTA AAGGCTTCTT
---------------------------------------------------------------------
1451 GACCTACTCG AGATGGGCGT GATTACGGAT TCACTGGCCG TCGTTTTACA
     CTGGATGAGC TCTACCCGCA CTAATGCCTA AGTGACCGGC AGCAAAATGT
---------------------------------------------------------------------
1501 ACGTCGTGAC TGGGAAAACC CTGGCGTTAC CCAACTTAAT CGCCTTGCAG
     TGCAGCACTG ACCCTTTTGG GACCGCAATG GGTTGAATTA GCGGAACGTC
---------------------------------------------------------------------
1551 CACATCCCCC TTTCGCCAGC TGGCGTAATA GCGAAGAGGC CCGCACCGAT
     GTGTAGGGGG AAAGCGGTCG ACCGCATTAT CGCTTCTCCG GGCGTGGCTA
---------------------------------------------------------------------
1601 CGCCCTTCCC AACAGTTACG CAGCCTGAAT GGCGAATGGC GCTTTGCCTG
     GCGGGAAGGG TTGTCAATGC GTCGGACTTA CCGCTTACCG CGAAACGGAC
---------------------------------------------------------------------
1651 GTTTCCGGCA CCAGAAGCGG TGCCGGAAAG CTGGCTGGAG TGCGATCTTC
     CAAAGGCCGT GGTCTTCGCC ACGGCCTTTC GACCGACCTC ACGCTAGAAG
---------------------------------------------------------------------
1701 CTGAGGCCGA TACTGTCGTC GTCCCCTCAA ACTGGCAGAT GCACGGTTAC
     GACTCCGGCT ATGACAGCAG CAGGGGAGTT TGACCGTCTA CGTGCCAATG
---------------------------------------------------------------------
1751 GATGCGCCCA TCTACACCAA CGTGACCTAT CCCATTACGG TCAATCCGCC
     CTACGCGGGT AGATGTGGTT GCACTGGATA GGGTAATGCC AGTTAGGCGG
---------------------------------------------------------------------
1801 GTTTGTTCCC ACGGAGAATC CGACGGGTTG TTACTCGCTC ACATTTAATG
     CAAACAAGGG TGCCTCTTAG GCTGCCCAAC AATGAGCGAG TGTAAATTAC
---------------------------------------------------------------------
1851 TTGATGAAAG CTGGCTACAG GAAGGCCAGA CGCGAATTAT TTTTGATGGC
     AACTACTTTC GACCGATGTC CTTCCGGTCT GCGCTTAATA AAAACTACCG
---------------------------------------------------------------------
```

```
1901  GTTAACTCGG CGTTTCATCT GTGGTGCAAC GGGCGCTGGG TCGGTTACGG
      CAATTGAGCC GCAAAGTAGA CACCACGTTG CCCGCGACCC AGCCAATGCC
-----------------------------------------------------------------------
1951  CCAGGACAGT CGTTTGCCGT CTGAATTTGA CCTGAGCGCA TTTTTACGCG
      GGTCCTGTCA GCAAACGGCA GACTTAAACT GGACTCGCGT AAAAATGCGC
-----------------------------------------------------------------------
2001  CCGGAGAAAA CCGCCTCGCG GTGATGGTGC TGCGCTGGAG TGACGGCAGT
      GGCCTCTTTT GGCGGAGCGC CACTACCACG ACGCGACCTC ACTGCCGTCA
-----------------------------------------------------------------------
2051  TATCTGGAAG ATCAGGATAT GTGGCGGATG AGCGGCATTT TCCGTGACGT
      ATAGACCTTC TAGTCCTATA CACCGCCTAC TCGCCGTAAA AGGCACTGCA
-----------------------------------------------------------------------
2101  CTCGTTGCTG CATAAACCGA CTACACAAAT CAGCGATTTC CATGTTGCCA
      GAGCAACGAC GTATTTGGCT GATGTGTTTA GTCGCTAAAG GTACAACGGT
-----------------------------------------------------------------------
2151  CTCGCTTTAA TGATGATTTC AGCCGCGCTG TACTGGAGGC TGAAGTTCAG
      GAGCGAAATT ACTACTAAAG TCGGCGCGAC ATGACCTCCG ACTTCAAGTC
-----------------------------------------------------------------------
2201  ATGTGCGGCG AGTTGCGTGA CTACCTACGG GTAACAGTTT CTTTATGGCA
      TACACGCCGC TCAACGCACT GATGGATGCC CATTGTCAAA GAAATACCGT
-----------------------------------------------------------------------
2251  GGGTGAAACG CAGGTCGCCA GCGGCACCGC GCCTTTCGGC GGTGAAATTA
      CCCACTTTGC GTCCAGCGGT CGCCGTGGCG CGGAAAGCCG CCACTTTAAT
-----------------------------------------------------------------------
2301  TCGATGAGCG TGGTGGTTAT GCCGATCGCG TCACACTACG TCTGAACGTC
      AGCTACTCGC ACCACCAATA CGGCTAGCGC AGTGTGATGC AGACTTGCAG
-----------------------------------------------------------------------
2351  GAAAACCCGA AACTGTGGAG CGCCGAAATC CCGAATCTCT ATCGTGCGGT
      CTTTTGGGCT TTGACACCTC GCGGCTTTAG GGCTTAGAGA TAGCACGCCA
-----------------------------------------------------------------------
2401  GGTTGAACTG CACACCGCCG ACGGCACGCT GATTGAAGCA GAAGCCTGCG
      CCAACTTGAC GTGTGGCGGC TGCCGTGCGA CTAACTTCGT CTTCGGACGC
-----------------------------------------------------------------------
2451  ATGTCGGTTT CCGCGAGGTG CGGATTGAAA ATGGTCTGCT GCTGCTGAAC
      TACAGCCAAA GGCGCTCCAC GCCTAACTTT TACCAGACGA CGACGACTTG
-----------------------------------------------------------------------
2501  GGCAAGCCGT TGCTGATTCG AGGCGTTAAC CGTCACGAGC ATCATCCTCT
      CCGTTCGGCA ACGACTAAGC TCCGCAATTG GCAGTGCTCG TAGTAGGAGA
-----------------------------------------------------------------------
2551  GCATGGTCAG GTCATGGATG AGCAGACGAT GGTGCAGGAT ATCCTGCTGA
      CGTACCAGTC CAGTACCTAC TCGTCTGCTA CCACGTCCTA TAGGACGACT
-----------------------------------------------------------------------
2601  TGAAGCAGAA CAACTTTAAC GCCGTGCGCT GTTCGCATTA TCCGAACCAT
      ACTTCGTCTT GTTGAAATTG CGGCACGCGA CAAGCGTAAT AGGCTTGGTA
-----------------------------------------------------------------------
2651  CCGCTGTGGT ACACGCTGTG CGACCGCTAC GGCCTGTATG TGGTGGATGA
      GGCGACACCA TGTGCGACAC GCTGGCGATG CCGGACATAC ACCACCTACT
-----------------------------------------------------------------------
2701  AGCCAATATT GAAACCCACG GCATGGTGCC AATGAATCGT CTGACCGATG
      TCGGTTATAA CTTTGGGTGC CGTACCACGG TTACTTAGCA GACTGGCTAC
-----------------------------------------------------------------------
2751  ATCCGCGCTG GCTACCGGCG ATGAGCGAAC GCGTAACGCG AATGGTGCAG
      TAGGCGCGAC CGATGGCCGC TACTCGCTTG CGCATTGCGC TTACCACGTC
-----------------------------------------------------------------------
2801  CGCGATCGTA ATCACCCGAG TGTGATCATC TGGTCGCTGG GGAATGAATC
      GCGCTAGCAT TAGTGGGCTC ACACTAGTAG ACCAGCGACC CCTTACTTAG
-----------------------------------------------------------------------
```

```
2851  AGGCCACGGC GCTAATCACG ACGCGCTGTA TCGCTGGATC AAATCTGTCG
      TCCGGTGCCG CGATTAGTGC TGCGCGACAT AGCGACCTAG TTTAGACAGC

2901  ATCCTTCCCG CCCGGTGCAG TATGAAGGCG GCGGAGCCGA CACCACGGCC
      TAGGAAGGGC GGGCCACGTC ATACTTCCGC CGCCTCGGCT GTGGTGCCGG

2951  ACCGATATTA TTTGCCCGAT GTACGCGCGC GTGGATGAAG ACCAGCCCTT
      TGGCTATAAT AAACGGGCTA CATGCGCGCG CACCTACTTC TGGTCGGGAA

3001  CCCGGCTGTG CCGAAATGGT CCATCAAAAA ATGGCTTTCG CTACCTGGAG
      GGGCCGACAC GGCTTTACCA GGTAGTTTTT TACCGAAAGC GATGGACCTC

3051  AGACGCGCCC GCTGATCCTT TGCGAATACG CCCACGCGAT GGGTAACAGT
      TCTGCGCGGG CGACTAGGAA ACGCTTATGC GGGTGCGCTA CCCATTGTCA

3101  CTTGGCGGTT TCGCTAAATA CTGGCAGGCG TTTCGTCAGT ATCCCCGTTT
      GAACCGCCAA AGCGATTTAT GACCGTCCGC AAAGCAGTCA TAGGGGCAAA

3151  ACAGGGCGGC TTCGTCTGGG ACTGGGTGGA TCAGTCGCTG ATTAAATATG
      TGTCCCGCCG AAGCAGACCC TGACCCACCT AGTCAGCGAC TAATTTATAC

3201  ATGAAAACGG CAACCCGTGG TCGGCTTACG GCGGTGATTT TGGCGATACG
      TACTTTTGCC GTTGGGCACC AGCCGAATGC CGCCACTAAA ACCGCTATGC

3251  CCGAACGATC GCCAGTTCTG TATGAACGGT CTGGTCTTTG CCGACCGCAC
      GGCTTGCTAG CGGTCAAGAC ATACTTGCCA GACCAGAAAC GGCTGGCGTG

3301  GCCGCATCCA GCGCTGACGG AAGCAAAACA CCAGCAGCAG TTTTTCCAGT
      CGGCGTAGGT CGCGACTGCC TTCGTTTTGT GGTCGTCGTC AAAAAGGTCA

3351  TCCGTTTATC CGGGCAAACC ATCGAAGTGA CCAGCGAATA CCTGTTCCGT
      AGGCAAATAG GCCCGTTTGG TAGCTTCACT GGTCGCTTAT GGACAAGGCA

3401  CATAGCGATA ACGAGCTCCT GCACTGGATG GTGGCGCTGG ATGGTAAGCC
      GTATCGCTAT TGCTCGAGGA CGTGACCTAC CACCGCGACC TACCATTCGG

3451  GCTGGCAAGC GGTGAAGTGC CTCTGGATGT CGCTCCACAA GGTAAACAGT
      CGACCGTTCG CCACTTCACG GAGACCTACA GCGAGGTGTT CCATTTGTCA

3501  TGATTGAACT GCCTGAACTA CCGCAGCCGG AGAGCGCCGG GCAACTCTGG
      ACTAACTTGA CGGACTTGAT GGCGTCGGCC TCTCGCGGCC CGTTGAGACC

3551  CTCACAGTAC GCGTAGTGCA ACCGAACGCG ACCGCATGGT CAGAAGCCGG
      GAGTGTCATG CGCATCACGT TGGCTTGCGC TGGCGTACCA GTCTTCGGCC

3601  GCACATCAGC GCCTGGCAGC AGTGGCGTCT GGCGGAAAAC CTCAGTGTGA
      CGTGTAGTCG CGGACCGTCG TCACCGCAGA CCGCCTTTTG GAGTCACACT

3651  CGCTCCCCGC CGCGTCCCAC GCCATCCCGC ATCTGACCAC CAGCGAAATG
      GCGAGGGGCG GCGCAGGGTG CGGTAGGGCG TAGACTGGTG GTCGCTTTAC

3701  GATTTTTGCA TCGAGCTGGG TAATAAGCGT TGGCAATTTA ACCGCCAGTC
      CTAAAAACGT AGCTCGACCC ATTATTCGCA ACCGTTAAAT TGGCGGTCAG

3751  AGGCTTTCTT TCACAGATGT GGATTGGCGA TAAAAAACAA CTGCTGACGC
      TCCGAAAGAA AGTGTCTACA CCTAACCGCT ATTTTTTGTT GACGACTGCG
```

```
3801  CGCTGCGCGA TCAGTTCACC CGTGTCGATA GATCTGGAGG TGGTGGCAGC
      GCGACGCGCT AGTCAAGTGG GCACAGCTAT CTAGACCTCC ACCACCGTCG

3851  AGGCCTTGGC GCGCCGGATC CTTAATTAAC AATTGACCGG TAATAATAGG
      TCCGGAACCG CGCGGCCTAG GAATTAATTG TTAACTGGCC ATTATTATCC

3901  TAGATAAGTG ACTGATTAGA TGCATTTCGA CTAGATCCCT CGACCAATTC
      ATCTATTCAC TGACTAATCT ACGTAAAGCT GATCTAGGGA GCTGGTTAAG

3951  CGGTTATTTT CCACCATATT GCCGTCTTTT GGCAATGTGA GGGCCCGGAA
      GCCAATAAAA GGTGGTATAA CGGCAGAAAA CCGTTACACT CCCGGGCCTT

4001  ACCTGGCCCT GTCTTCTTGA CGAGCATTCC TAGGGGTCTT TCCCCTCTCG
      TGGACCGGGA CAGAAGAACT GCTCGTAAGG ATCCCCAGAA AGGGGAGAGC

4051  CCAAAGGAAT GCAAGGTCTG TTGAATGTCG TGAAGGAAGC AGTTCCTCTG
      GGTTTCCTTA CGTTCCAGAC AACTTACAGC ACTTCCTTCG TCAAGGAGAC

4101  GAAGCTTCTT GAAGACAAAC AACGTCTGTA GCGACCCTTT GCAGGCAGCG
      CTTCGAAGAA CTTCTGTTTG TTGCAGACAT CGCTGGGAAA CGTCCGTCGC

4151  GAACCCCCCA CCTGGCGACA GGTGCCTCTG CGGCCAAAAG CCACGTGTAT
      CTTGGGGGGT GGACCGCTGT CCACGGAGAC GCCGGTTTTC GGTGCACATA

4201  AAGATACACC TGCAAAGGCG GCACAACCCC AGTGCCACGT TGTGAGTTGG
      TTCTATGTGG ACGTTTCCGC CGTGTTGGGG TCACGGTGCA ACACTCAACC

4251  ATAGTTGTGG AAAGAGTCAA ATGGCTCTCC TCAAGCGTAT TCAACAAGGG
      TATCAACACC TTTCTCAGTT TACCGAGAGG AGTTCGCATA AGTTGTTCCC

4301  GCTGAAGGAT GCCCAGAAGG TACCCCATTG TATGGGATCT GATCTGGGGC
      CGACTTCCTA CGGGTCTTCC ATGGGGTAAC ATACCCTAGA CTAGACCCCG

4351  CTCGGTGCAC ATGCTTTACA TGTGTTTAGT CGAGGTTAAA AAACGTCTAG
      GAGCCACGTG TACGAAATGT ACACAAATCA GCTCCAATTT TTTGCAGATC

4401  GCCCCCCGAA CCACGGGGAC GTGGTTTTCC TTTGAAAAAC ACGATGATAA
      CGGGGGGCTT GGTGCCCCTG CACCAAAAGG AAACTTTTTG TGCTACTATT

4451  TACCATGAAA AAGCCTGAAC TCACCGCGAC GTCTGTCGAG AAGTTTCTGA
      ATGGTACTTT TTCGGACTTG AGTGGCGCTG CAGACAGCTC TTCAAAGACT

4501  TCGAAAAGTT CGACAGCGTC TCCGACCTGA TGCAGCTCTC GGAGGGCGAA
      AGCTTTTCAA GCTGTCGCAG AGGCTGGACT ACGTCGAGAG CCTCCCGCTT

4551  GAATCTCGTG CTTTCAGCTT CGATGTAGGA GGGCGTGGAT ATGTCCTGCG
      CTTAGAGCAC GAAAGTCGAA GCTACATCCT CCCGCACCTA TACAGGACGC

4601  GGTAAATAGC TGCGCCGATG GTTTCTACAA AGATCGTTAT GTTTATCGGC
      CCATTTATCG ACGCGGCTAC CAAAGATGTT TCTAGCAATA CAAATAGCCG

4651  ACTTTGCATC GGCCGCGCTC CCGATTCCGG AAGTGCTTGA CATTGGGGAA
      TGAAACGTAG CCGGCGCGAG GGCTAAGGCC TTCACGAACT GTAACCCCTT

4701  TTTAGCGAGA GCCTGACCTA TTGCATCTCC CGCCGTGCAC AGGGTGTCAC
      AAATCGCTCT CGGACTGGAT AACGTAGAGG GCGGCACGTG TCCCACAGTG
```

```
4751  GTTGCAAGAC CTGCCTGAAA CCGAACTGCC CGCTGTTCTG CAGCCGGTCG
      CAACGTTCTG GACGGACTTT GGCTTGACGG GCGACAAGAC GTCGGCCAGC
-------------------------------------------------------------------------
4801  CGGAGGCCAT GGATGCGATC GCTGCGGCCG ATCTTAGCCA GACGAGCGGG
      GCCTCCGGTA CCTACGCTAG CGACGCCGGC TAGAATCGGT CTGCTCGCCC
-------------------------------------------------------------------------
4851  TTCGGCCCAT TCGGACCGCA AGGAATCGGT CAATACACTA CATGGCGTGA
      AAGCCGGGTA AGCCTGGCGT TCCTTAGCCA GTTATGTGAT GTACCGCACT
-------------------------------------------------------------------------
4901  TTTCATATGC GCGATTGCTG ATCCCCATGT GTATCACTGG CAAACTGTGA
      AAAGTATACG CGCTAACGAC TAGGGGTACA CATAGTGACC GTTTGACACT
-------------------------------------------------------------------------
4951  TGGACGACAC CGTCAGTGCG TCCGTCGCGC AGGCTCTCGA TGAGCTGATG
      ACCTGCTGTG GCAGTCACGC AGGCAGCGCG TCCGAGAGCT ACTCGACTAC
-------------------------------------------------------------------------
5001  CTTTGGGCCG AGGACTGCCC CGAAGTCCGG CACCTCGTGC ACGCGGATTT
      GAAACCCGGC TCCTGACGGG GCTTCAGGCC GTGGAGCACG TGCGCCTAAA
-------------------------------------------------------------------------
5051  CGGCTCCAAC AATGTCCTGA CGGACAATGG CCGCATAACA GCGGTCATTG
      GCCGAGGTTG TTACAGGACT GCCTGTTACC GGCGTATTGT CGCCAGTAAC
-------------------------------------------------------------------------
5101  ACTGGAGCGA GGCGATGTTC GGGGATTCCC AATACGAGGT CGCCAACATC
      TGACCTCGCT CCGCTACAAG CCCCTAAGGG TTATGCTCCA GCGGTTGTAG
-------------------------------------------------------------------------
5151  TTCTTCTGGA GGCCGTGGTT GGCTTGTATG GAGCAGCAGA CGCGCTACTT
      AAGAAGACCT CCGGCACCAA CCGAACATAC CTCGTCGTCT GCGCGATGAA
-------------------------------------------------------------------------
5201  CGAGCGGAGG CATCCGGAGC TTGCAGGATC GCCGCGGCTC CGGGCGTATA
      GCTCGCCTCC GTAGGCCTCG AACGTCCTAG CGGCGCCGAG GCCCGCATAT
-------------------------------------------------------------------------
5251  TGCTCCGCAT TGGTCTTGAC CAACTCTATC AGAGCTTGGT TGACGGCAAT
      ACGAGGCGTA ACCAGAACTG GTTGAGATAG TCTCGAACCA ACTGCCGTTA
-------------------------------------------------------------------------
5301  TTCGATGATG CAGCTTGGGC GCAGGGTCGA TGCGACGCAA TCGTCCGATC
      AAGCTACTAC GTCGAACCCG CGTCCCAGCT ACGCTGCGTT AGCAGGCTAG
-------------------------------------------------------------------------
5351  CGGAGCCGGG ACTGTCGGGC GTACACAAAT CGCCCGCAGA AGCGCGGCCG
      GCCTCGGCCC TGACAGCCCG CATGTGTTTA GCGGGCGTCT TCGCGCCGGC
-------------------------------------------------------------------------
5401  TCTGGACCGA TGGCTGTGTA GAAGTACTCG CCGATAGTGG AAACCGACGC
      AGACCTGGCT ACCGACACAT CTTCATGAGC GGCTATCACC TTTGGCTGCG
-------------------------------------------------------------------------
5451  CCCAGCACTC GTCCGAGGGC AAAGGAATAG AGTAGATGCC GACCGGGATC
      GGGTCGTGAG CAGGCTCCCG TTTCCTTATC TCATCTACGG CTGGCCCTAG
-------------------------------------------------------------------------
5501  TATCGATAAA ATAAAAGATT TTATTTAGTC TCCAGAAAAA GGGGGGAATG
      ATAGCTATTT TATTTTCTAA AATAAATCAG AGGTCTTTTT CCCCCCTTAC
-------------------------------------------------------------------------
5551  AAAGACCCCA CCTGTAGGTT TGGCAAGCTA GCTTAAGTAA CGCCATTTTG
      TTTCTGGGGT GGACATCCAA ACCGTTCGAT CGAATTCATT GCGGTAAAAC
-------------------------------------------------------------------------
5601  CAAGGCATGG AAAAATACAT AACTGAGAAT AGAGAAGTTC AGATCAAGGT
      GTTCCGTACC TTTTTATGTA TTGACTCTTA TCTCTTCAAG TCTAGTTCCA
-------------------------------------------------------------------------
5651  CAGGAACAGA TGGAACAGCT GAATATGGGC CAAACAGGAT ATCTGTGGTA
      GTCCTTGTCT ACCTTGTCGA CTTATACCCG GTTTGTCCTA TAGACACCAT
-------------------------------------------------------------------------
```

```
5701  AGCAGTTCCT GCCCCGGCTC AGGGCCAAGA ACAGATGGAA CAGCTGAATA
      TCGTCAAGGA CGGGGCCGAG TCCCGGTTCT TGTCTACCTT GTCGACTTAT
----------------------------------------------------------------------
5751  TGGGCCAAAC AGGATATCTG TGGTAAGCAG TTCCTGCCCC GGCTCAGGGC
      ACCCGGTTTG TCCTATAGAC ACCATTCGTC AAGGACGGGG CCGAGTCCCG
----------------------------------------------------------------------
5801  CAAGAACAGA TGGTCCCCAG ATGCGGTCCA GCCCTCAGCA GTTTCTAGAG
      GTTCTTGTCT ACCAGGGGTC TACGCCAGGT CGGGAGTCGT CAAAGATCTC
----------------------------------------------------------------------
5851  AACCATCAGA TGTTTCCAGG GTGCCCCAAG GACCTGAAAT GACCCTGTGC
      TTGGTAGTCT ACAAAGGTCC CACGGGGTTC CTGGACTTTA CTGGGACACG
----------------------------------------------------------------------
5901  CTTATTTGAA CTAACCAATC AGTTCGCTTC TCGCTTCTGT TCGCGCGCTT
      GAATAAACTT GATTGGTTAG TCAAGCGAAG AGCGAAGACA AGCGCGCGAA
----------------------------------------------------------------------
5951  CTGCTCCCCG AGCTCAATAA AAGAGCCCAC AACCCCTCAC TCGGGGCGCC
      GACGAGGGGC TCGAGTTATT TTCTCGGGTG TTGGGGAGTG AGCCCCGCGG
----------------------------------------------------------------------
6001  AGTCCTCCGA TTGACTGAGT CGCCCGGGTA CCCGTGTATC CAATAAACCC
      TCAGGAGGCT AACTGACTCA GCGGGCCCAT GGGCACATAG GTTATTTGGG
----------------------------------------------------------------------
6051  TCTTGCAGTT GCATCCGACT TGTGGTCTCG CTGTTCCTTG GGAGGGTCTC
      AGAACGTCAA CGTAGGCTGA ACACCAGAGC GACAAGGAAC CCTCCCAGAG
----------------------------------------------------------------------
6101  CTCTGAGTGA TTGACTACCC GTCAGCGGGG GTCTTTCATT CATGCAGCAT
      GAGACTCACT AACTGATGGG CAGTCGCCCC CAGAAAGTAA GTACGTCGTA
----------------------------------------------------------------------
6151  GTATCAAAAT TAATTTGGTT TTTTTTCTTA AGTATTTACA TTAAATGGCC
      CATAGTTTTA ATTAAACCAA AAAAAAGAAT TCATAAATGT AATTTACCGG
----------------------------------------------------------------------
6201  ATAGTTGCAT TAATGAATCG GCCAACGCGC GGGGAGAGGC GGTTTGCGTA
      TATCAACGTA ATTACTTAGC CGGTTGCGCG CCCCTCTCCG CCAAACGCAT
----------------------------------------------------------------------
6251  TTGGCGCTCT TCCGCTTCCT CGCTCACTGA CTCGCTGCGC TCGGTCGTTC
      AACCGCGAGA AGGCGAAGGA GCGAGTGACT GAGCGACGCG AGCCAGCAAG
----------------------------------------------------------------------
6301  GGCTGCGGCG AGCGGTATCA GCTCACTCAA AGGCGGTAAT ACGGTTATCC
      CCGACGCCGC TCGCCATAGT CGAGTGAGTT TCCGCCATTA TGCCAATAGG
----------------------------------------------------------------------
6351  ACAGAATCAG GGGATAACGC AGGAAAGAAC ATGTGAGCAA AAGGCCAGCA
      TGTCTTAGTC CCCTATTGCG TCCTTTCTTG TACACTCGTT TTCCGGTCGT
----------------------------------------------------------------------
6401  AAAGGCCAGG AACCGTAAAA AGGCCGCGTT GCTGGCGTTT TTCCATAGGC
      TTTCCGGTCC TTGGCATTTT TCCGGCGCAA CGACCGCAAA AAGGTATCCG
----------------------------------------------------------------------
6451  TCCGCCCCCC TGACGAGCAT CACAAAAATC GACGCTCAAG TCAGAGGTGG
      AGGCGGGGGG ACTGCTCGTA GTGTTTTTAG CTGCGAGTTC AGTCTCCACC
----------------------------------------------------------------------
6501  CGAAACCCGA CAGGACTATA AAGATACCAG GCGTTTCCCC CTGGAAGCTC
      GCTTTGGGCT GTCCTGATAT TTCTATGGTC CGCAAAGGGG GACCTTCGAG
----------------------------------------------------------------------
6551  CCTCGTGCGC TCTCCTGTTC CGACCCTGCC GCTTACCGGA TACCTGTCCG
      GGAGCACGCG AGAGGACAAG GCTGGGACGG CGAATGGCCT ATGGACAGGC
----------------------------------------------------------------------
6601  CCTTTCTCCC TTCGGGAAGC GTGGCGCTTT CTCATAGCTC ACGCTGTAGG
      GGAAAGAGGG AAGCCCTTCG CACCGCGAAA GAGTATCGAG TGCGACATCC
----------------------------------------------------------------------
```

```
6651  TATCTCAGTT CGGTGTAGGT CGTTCGCTCC AAGCTGGGCT GTGTGCACGA
      ATAGAGTCAA GCCACATCCA GCAAGCGAGG TTCGACCCGA CACACGTGCT
------------------------------------------------------------------------
6701  ACCCCCCGTT CAGCCCGACC GCTGCGCCTT ATCCGGTAAC TATCGTCTTG
      TGGGGGGCAA GTCGGGCTGG CGACGCGGAA TAGGCCATTG ATAGCAGAAC
------------------------------------------------------------------------
6751  AGTCCAACCC GGTAAGACAC GACTTATCGC CACTGGCAGC AGCCACTGGT
      TCAGGTTGGG CCATTCTGTG CTGAATAGCG GTGACCGTCG TCGGTGACCA
------------------------------------------------------------------------
6801  AACAGGATTA GCAGAGCGAG GTATGTAGGC GGTGCTACAG AGTTCTTGAA
      TTGTCCTAAT CGTCTCGCTC CATACATCCG CCACGATGTC TCAAGAACTT
------------------------------------------------------------------------
6851  GTGGTGGCCT AACTACGGCT ACACTAGAAG AACAGTATTT GGTATCTGCG
      CACCACCGGA TTGATGCCGA TGTGATCTTC TTGTCATAAA CCATAGACGC
------------------------------------------------------------------------
6901  CTCTGCTGAA GCCAGTTACC TTCGGAAAAA GAGTTGGTAG CTCTTGATCC
      GAGACGACTT CGGTCAATGG AAGCCTTTTT CTCAACCATC GAGAACTAGG
------------------------------------------------------------------------
6951  GGCAAACAAA CCACCGCTGG TAGCGGTGGT TTTTTTGTTT GCAAGCAGCA
      CCGTTTGTTT GGTGGCGACC ATCGCCACCA AAAAAACAAA CGTTCGTCGT
------------------------------------------------------------------------
7001  GATTACGCGC AGAAAAAAAG GATCTCAAGA AGATCCTTTG ATCTTTTCTA
      CTAATGCGCG TCTTTTTTTC CTAGAGTTCT TCTAGGAAAC TAGAAAAGAT
------------------------------------------------------------------------
7051  CGGGGTCTGA CGCTCAGTGG AACGAAAACT CACGTTAAGG GATTTTGGTC
      GCCCCAGACT GCGAGTCACC TTGCTTTTGA GTGCAATTCC CTAAAACCAG
------------------------------------------------------------------------
7101  ATGAGATTAT CAAAAAGGAT CTTCACCTAG ATCCTTTTGC GGCCGCAAAT
      TACTCTAATA GTTTTTCCTA GAAGTGGATC TAGGAAAACG CCGGCGTTTA
------------------------------------------------------------------------
7151  CAATCTAAAG TATATATGAG TAAACTTGGT CTGACAGTTA CCAATGCTTA
      GTTAGATTTC ATATATACTC ATTTGAACCA GACTGTCAAT GGTTACGAAT
------------------------------------------------------------------------
7201  ATCAGTGAGG CACCTATCTC AGCGATCTGT CTATTTCGTT CATCCATAGT
      TAGTCACTCC GTGGATAGAG TCGCTAGACA GATAAAGCAA GTAGGTATCA
------------------------------------------------------------------------
7251  TGCCTGACTC CCCGTCGTGT AGATAACTAC GATACGGGAG GGCTTACCAT
      ACGGACTGAG GGGCAGCACA TCTATTGATG CTATGCCCTC CCGAATGGTA
------------------------------------------------------------------------
7301  CTGGCCCCAG TGCTGCAATG ATACCGCGAG ACCCACGCTC ACCGGCTCCA
      GACCGGGGTC ACGACGTTAC TATGGCGCTC TGGGTGCGAG TGGCCGAGGT
------------------------------------------------------------------------
7351  GATTTATCAG CAATAAACCA GCCAGCCGGA AGGGCCGAGC GCAGAAGTGG
      CTAAATAGTC GTTATTTGGT CGGTCGGCCT TCCCGGCTCG CGTCTTCACC
------------------------------------------------------------------------
7401  TCCTGCAACT TTATCCGCCT CCATCCAGTC TATTAATTGT TGCCGGGAAG
      AGGACGTTGA AATAGGCGGA GGTAGGTCAG ATAATTAACA ACGGCCCTTC
------------------------------------------------------------------------
7451  CTAGAGTAAG TAGTTCGCCA GTTAATAGTT TGCGCAACGT TGTTGCCATT
      GATCTCATTC ATCAAGCGGT CAATTATCAA ACGCGTTGCA ACAACGGTAA
------------------------------------------------------------------------
7501  GCTACAGGCA TCGTGGTGTC ACGCTCGTCG TTTGGTATGG CTTCATTCAG
      CGATGTCCGT AGCACCACAG TGCGAGCAGC AAACCATACC GAAGTAAGTC
------------------------------------------------------------------------
7551  CTCCGGTTCC CAACGATCAA GGCGAGTTAC ATGATCCCCC ATGTTGTGCA
      GAGGCCAAGG GTTGCTAGTT CCGCTCAATG TACTAGGGGG TACAACACGT
------------------------------------------------------------------------
```

```
7601  AAAAAGCGGT TAGCTCCTTC GGTCCTCCGA TCGTTGTCAG AAGTAAGTTG
      TTTTTCGCCA ATCGAGGAAG CCAGGAGGCT AGCAACAGTC TTCATTCAAC
------------------------------------------------------------------------------
7651  GCCGCAGTGT TATCACTCAT GGTTATGGCA GCACTGCATA ATTCTCTTAC
      CGGCGTCACA ATAGTGAGTA CCAATACCGT CGTGACGTAT TAAGAGAATG
------------------------------------------------------------------------------
7701  TGTCATGCCA TCCGTAAGAT GCTTTTCTGT GACTGGTGAG TACTCAACCA
      ACAGTACGGT AGGCATTCTA CGAAAAGACA CTGACCACTC ATGAGTTGGT
------------------------------------------------------------------------------
7751  AGTCATTCTG AGAATAGTGT ATGCGGCGAC CGAGTTGCTC TTGCCCGGCG
      TCAGTAAGAC TCTTATCACA TACGCCGCTG GCTCAACGAG AACGGGCCGC
------------------------------------------------------------------------------
7801  TCAATACGGG ATAATACCGC GCCACATAGC AGAACTTTAA AAGTGCTCAT
      AGTTATGCCC TATTATGGCG CGGTGTATCG TCTTGAAATT TTCACGAGTA
------------------------------------------------------------------------------
7851  CATTGGAAAA CGTTCTTCGG GGCGAAAACT CTCAAGGATC TTACCGCTGT
      GTAACCTTTT GCAAGAAGCC CCGCTTTTGA GAGTTCCTAG AATGGCGACA
------------------------------------------------------------------------------
7901  TGAGATCCAG TTCGATGTAA CCCACTCGTG CACCCAACTG ATCTTCAGCA
      ACTCTAGGTC AAGCTACATT GGGTGAGCAC GTGGGTTGAC TAGAAGTCGT
------------------------------------------------------------------------------
7951  TCTTTTACTT TCACCAGCGT TTCTGGGTGA GCAAAAACAG GAAGGCAAAA
      AGAAAATGAA AGTGGTCGCA AAGACCCACT CGTTTTTGTC CTTCCGTTTT
------------------------------------------------------------------------------
8001  TGCCGCAAAA AAGGGAATAA GGGCGACACG GAAATGTTGA ATACTCATAC
      ACGGCGTTTT TTCCCTTATT CCCGCTGTGC CTTTACAACT TATGAGTATG
------------------------------------------------------------------------------
8051  TCTTCCTTTT TCAATATTAT TGAAGCATTT ATCAGGGTTA TTGTCTCATG
      AGAAGGAAAA AGTTATAATA ACTTCGTAAA TAGTCCCAAT AACAGAGTAC
------------------------------------------------------------------------------
8101  AGCGGATACA TATTTGAATG TATTTAGAAA AATAAACAAA TAGGGGTTCC
      TCGCCTATGT ATAAACTTAC ATAAATCTTT TTATTTGTTT ATCCCCAAGG
------------------------------------------------------------------------------
8151  GCGCACATTT C
      CGCGTGTAAA G
------------------------------------------------------------------------------
```

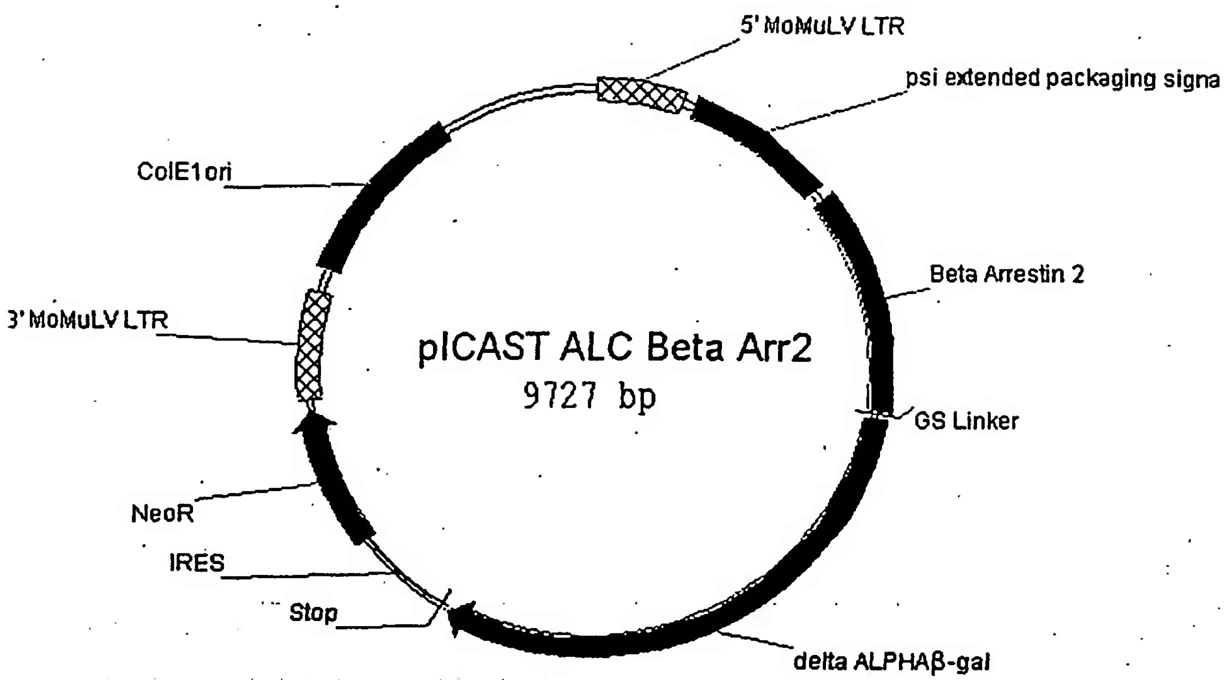

5' MoMuLV LTR
psi extended packaging signa
Beta Arrestin 2
GS Linker
delta ALPHAβ-gal
Stop
IRES
NeoR
3' MoMuLV LTR
ColE1 ori
pICAST ALC Beta Arr2
9727 bp

Figure 14

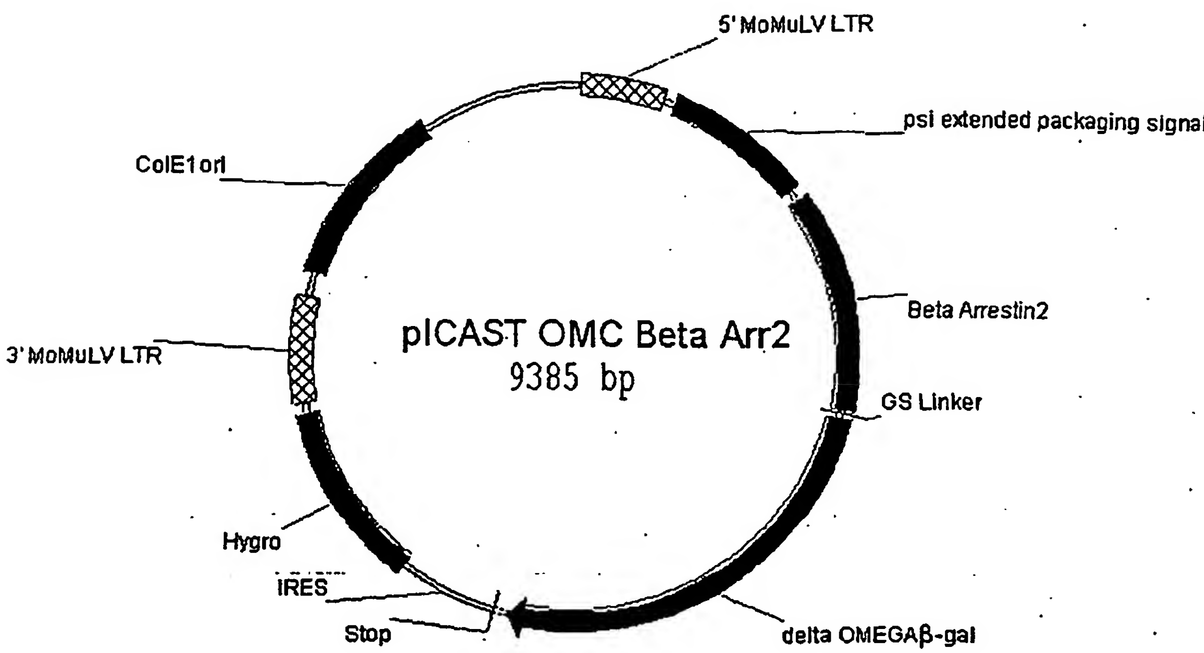

5' MoMuLV LTR
psi extended packaging signal
ColE1 ori
Beta Arrestin2
pICAST OMC Beta Arr2
9385 bp
3' MoMuLV LTR
GS Linker
Hygro
IRES
Stop
delta OMEGAβ-gal

Figure 15

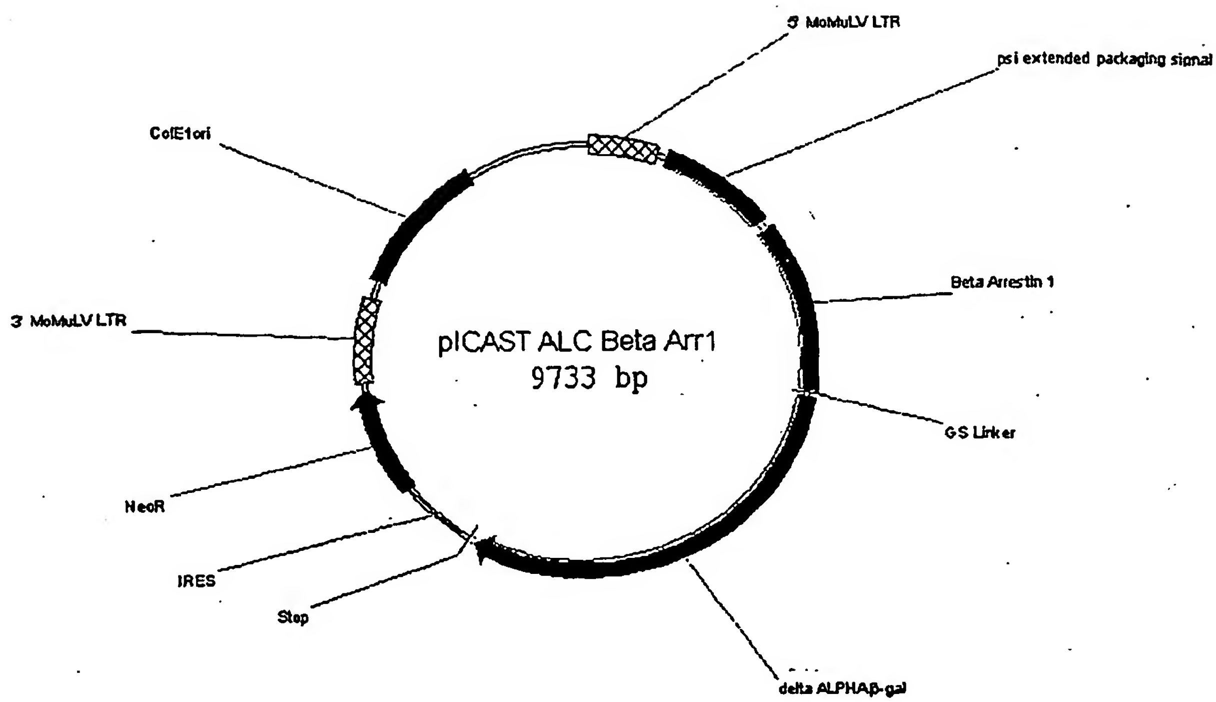

5' MoMuLV LTR
psi extended packaging signal
ColE1ori
Beta Arrestin 1
3' MoMuLV LTR
pICAST ALC Beta Arr1
9733 bp
GS Linker
NeoR
IRES
Stop
delta ALPHAβ-gal

Figure 16

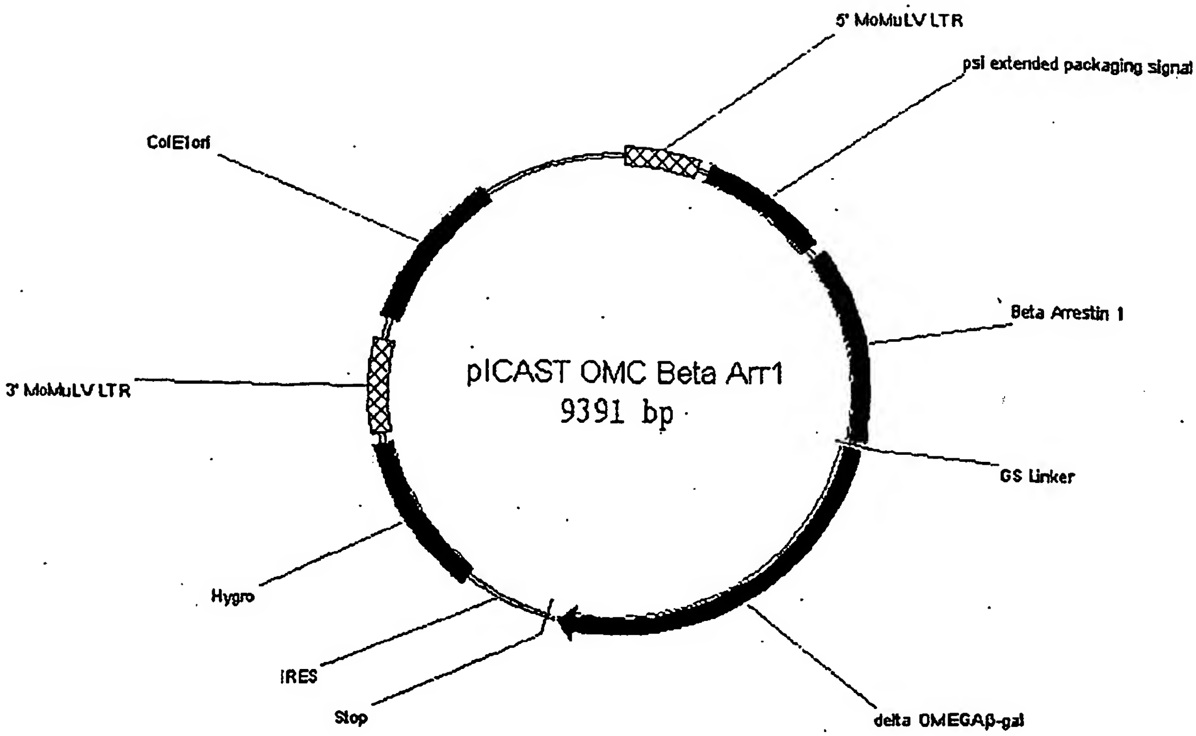

5' MoMuLV LTR
psi extended packaging signal
ColE1ori
Beta Arrestin 1
pICAST OMC Beta Arr1
9391 bp
3' MoMuLV LTR
GS Linker
Hygro
IRES
Stop
delta OMEGAβ-gal

Figure 17

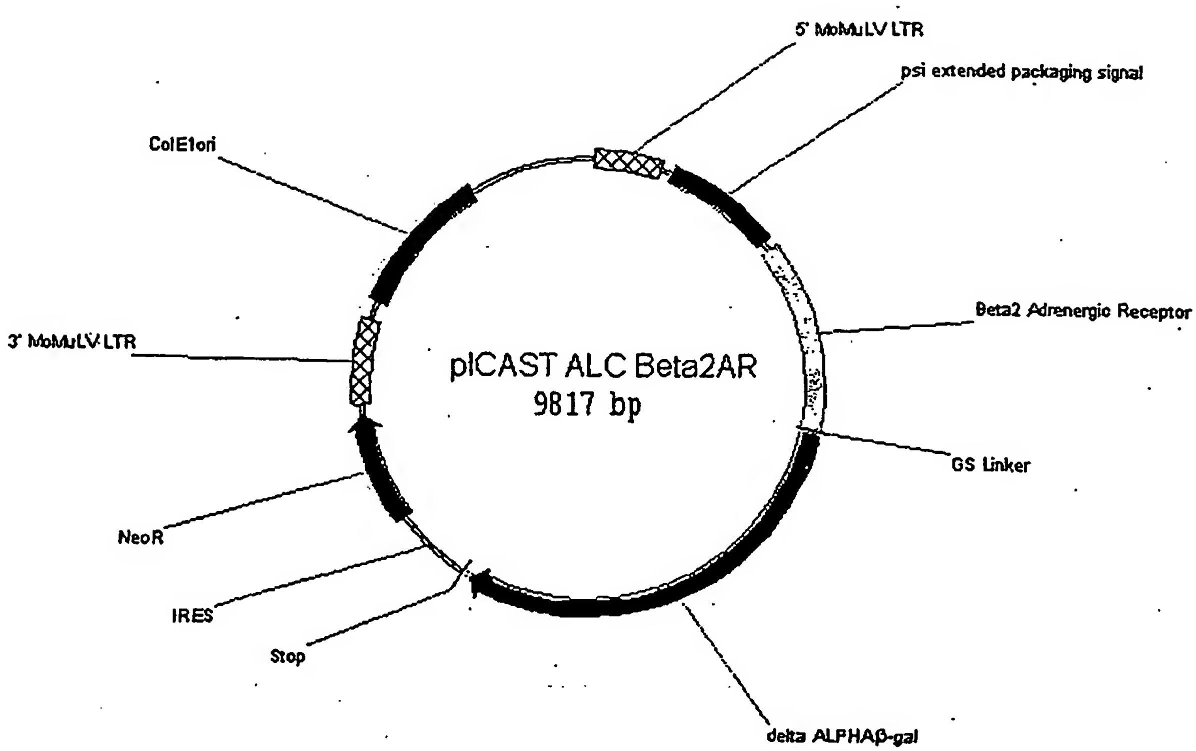

5' MoMuLV LTR
psi extended packaging signal
ColE1ori
Beta2 Adrenergic Receptor
3' MoMuLV LTR
pICAST ALC Beta2AR
9817 bp
GS Linker
NeoR
IRES
Stop
delta ALPHAβ-gal

Figure 18

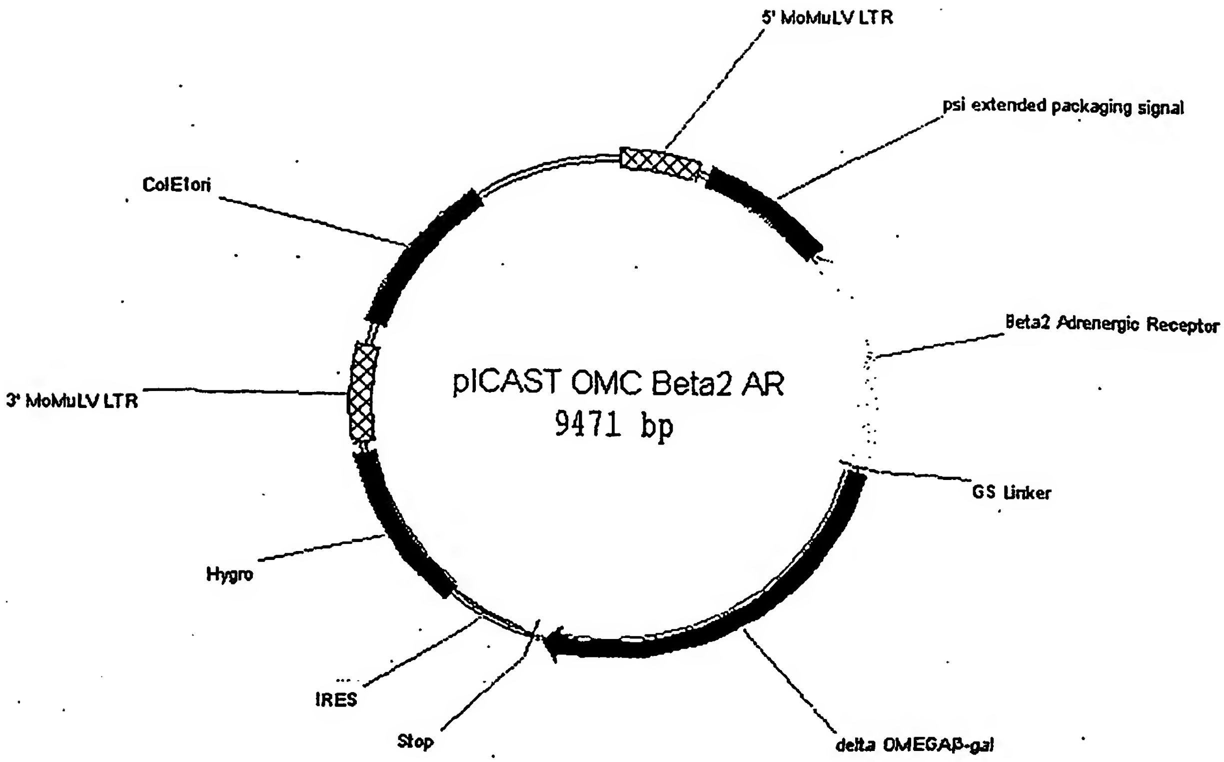

5' MoMuLV LTR
psi extended packaging signal
ColE1ori
Beta2 Adrenergic Receptor
pICAST OMC Beta2 AR
9471 bp
3' MoMuLV LTR
GS Linker
Hygro
IRES
Stop
delta OMEGAβ-gal

Figure 19

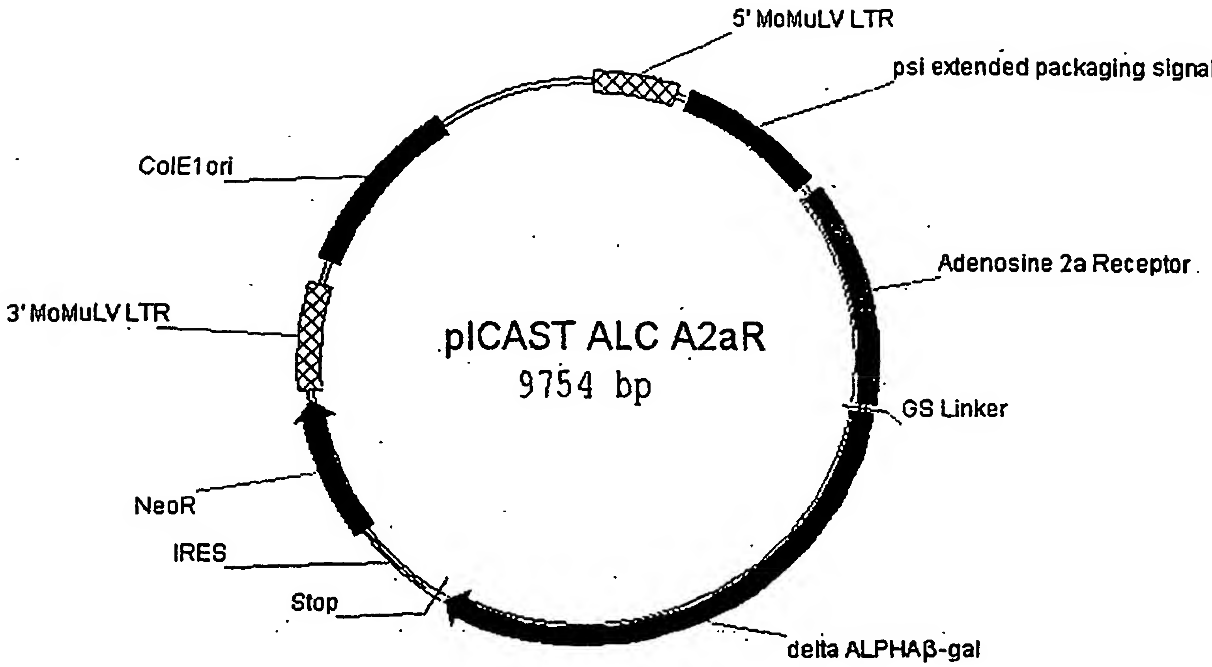

5' MoMuLV LTR
psi extended packaging signal
Adenosine 2a Receptor
GS Linker
delta ALPHAβ-gal
Stop
IRES
NeoR
3' MoMuLV LTR
ColE1 ori
pICAST ALC A2aR
9754 bp

Figure 20

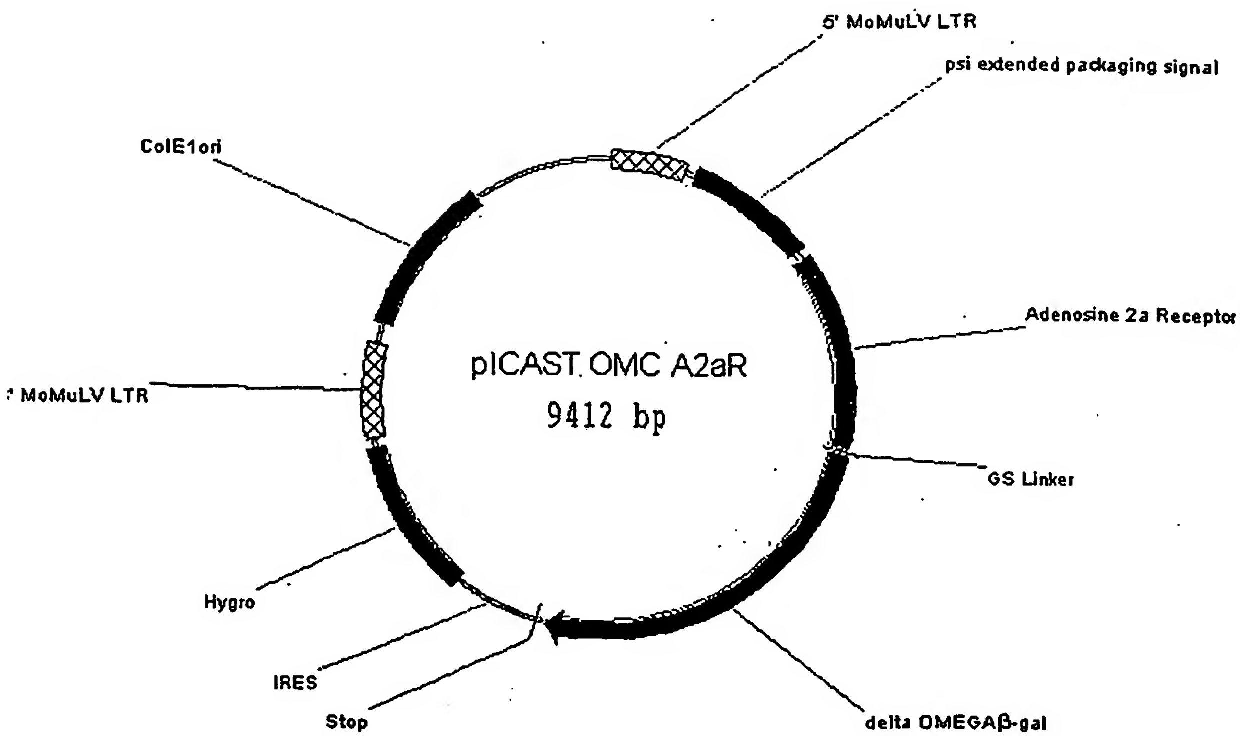

5' MoMuLV LTR
psi extended packaging signal
ColE1ori
Adenosine 2a Receptor
pICAST OMC A2aR
9412 bp
MoMuLV LTR
GS Linker
Hygro
IRES
Stop
delta OMEGAβ-gal

Figure 21

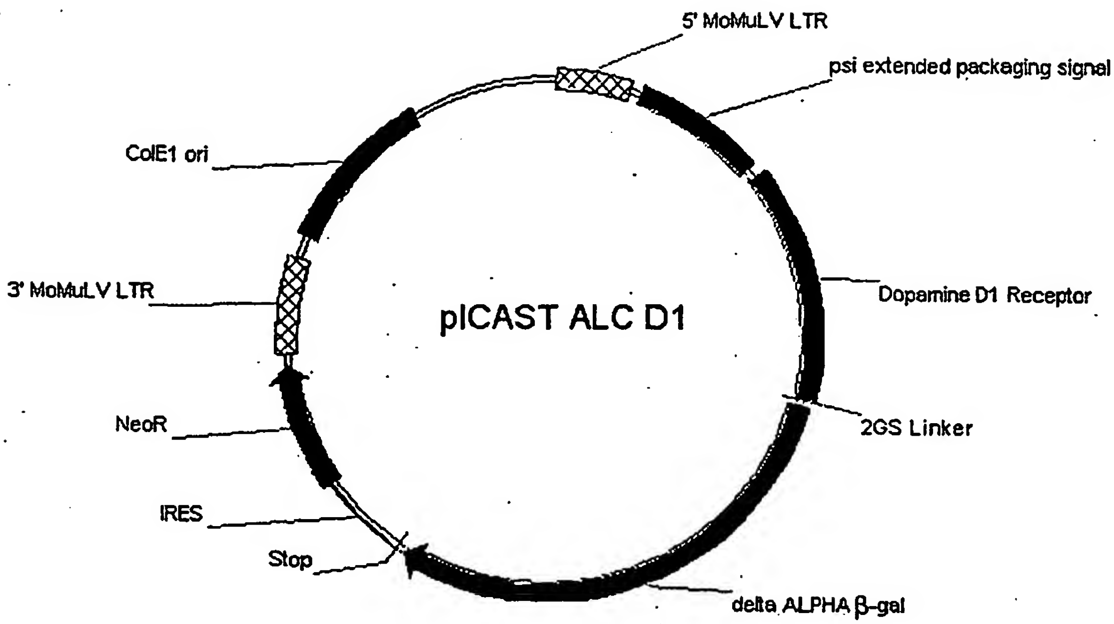

5' MoMuLV LTR
psi extended packaging signal
ColE1 ori
3' MoMuLV LTR
pICAST ALC D1
Dopamine D1 Receptor
2GS Linker
NeoR
IRES
Stop
delta ALPHA β-gal

Figure 22

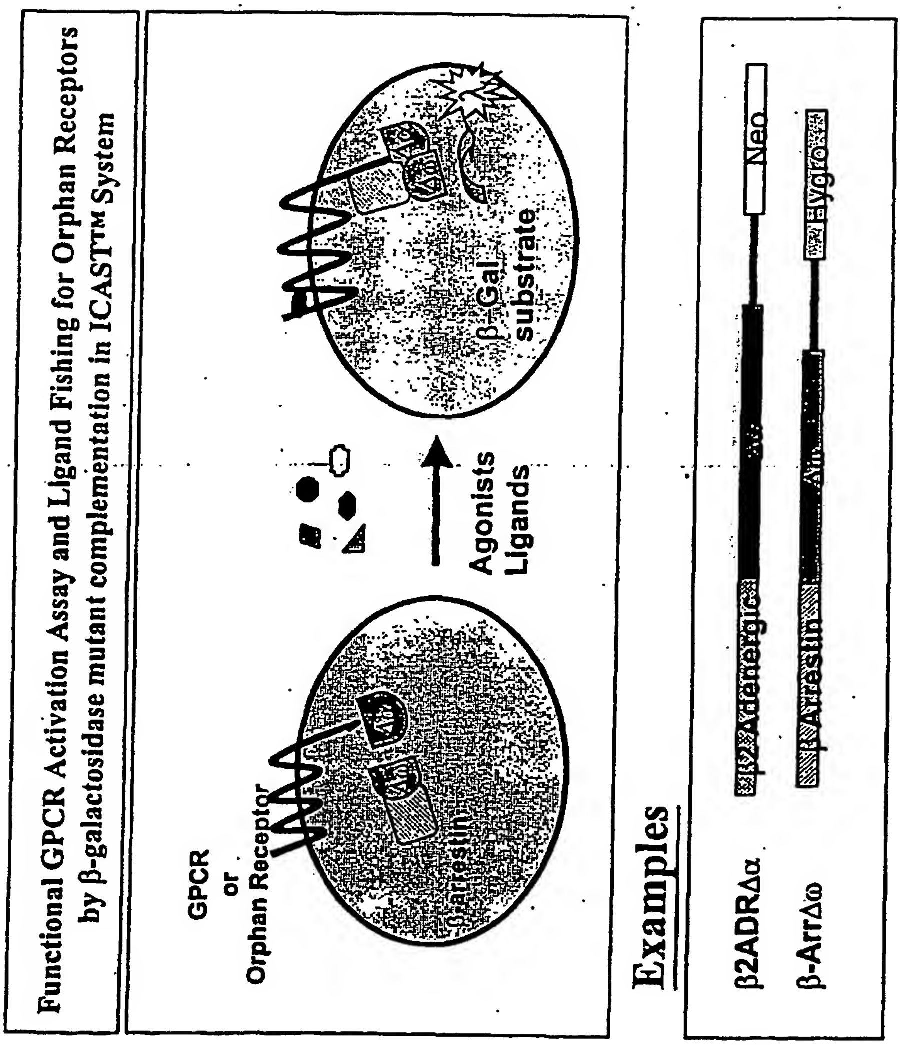
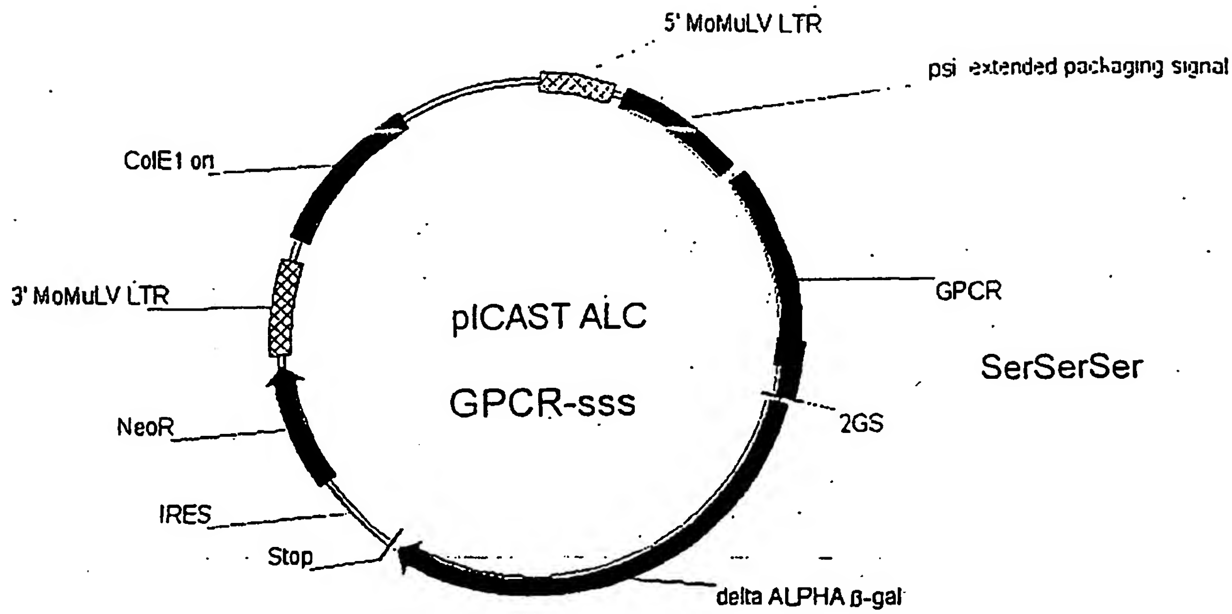
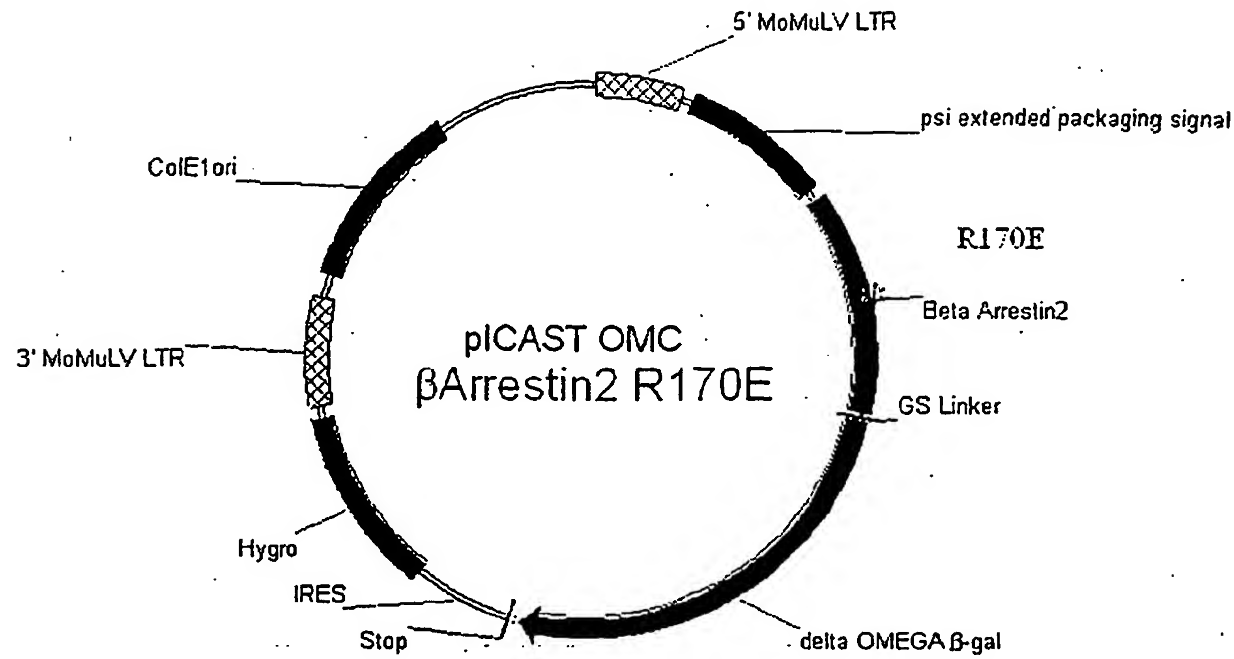

Functional GPCR Activation Assay and Ligand Fishing for Orphan Receptors
by β-galactosidase mutant complementation in ICAST™ System
GPCR
or
Orphan Receptor
β-arrestin
Agonists
Ligands
β-Gal
substrate
Examples
β2ADRΔα
β2 Adrenergic
Neo
β-ArrΔω
β-Arrestin
Hygro

Figure 23



Vector for Expression of a GPCR with inserted Seronine/Threonine amino acid sequences as a fusion with β-gal Δα.

FIGURE 24



Vector for Expression of mutant (R170E) β-arrestin2 as a fusion with β-gal Δω.

FIGURE 25

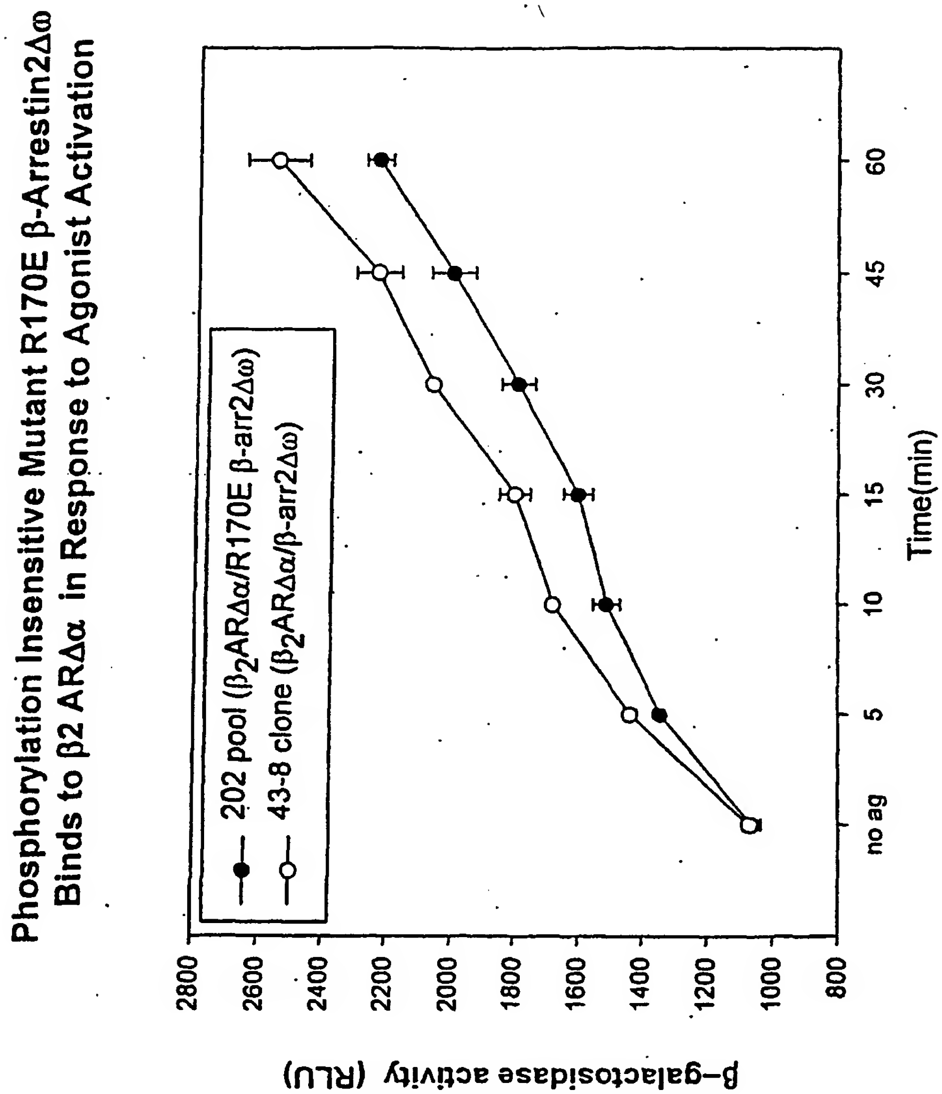
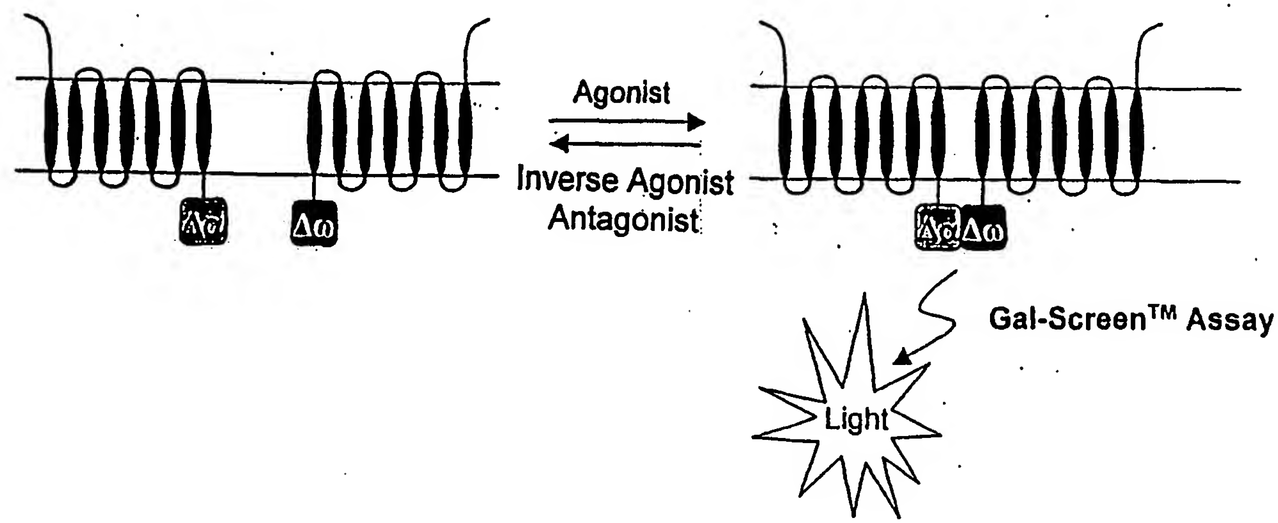

Phosphorylation Insensitive Mutant R170E β-Arrestin2Δω
Binds to β2 ARΔα in Response to Agonist Activation
202 pool (β2ARΔα/R170E β-arr2Δω)
43-8 clone (β2ARΔα/β-arr2Δω)
β-galactosidase activity (RLU)
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
no ag
5
10
15
30
45
60
Time(min)

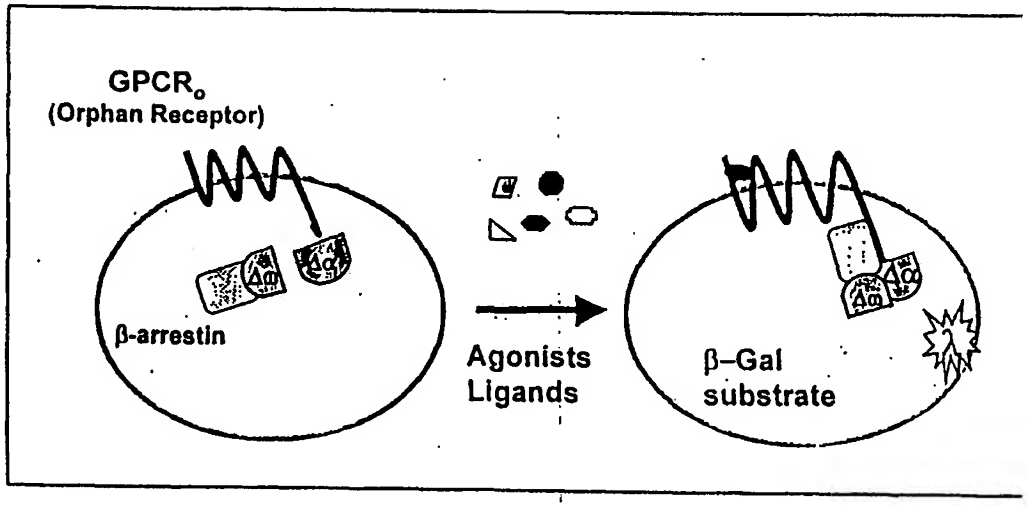
FIGURE 26

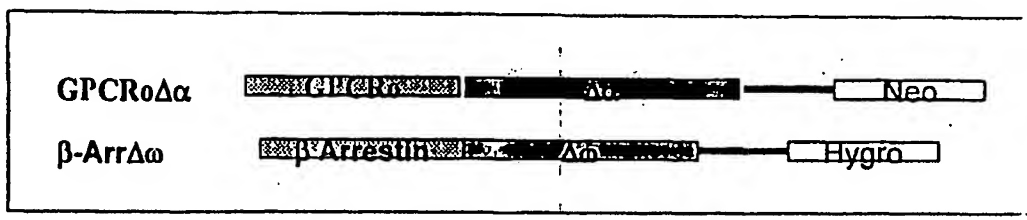


GPCR dimerization measured by β-gal complementation

FIGURE 27

Example-





Ligand Fishing for Orphan Receptors by β-galactosidase mutant complementation in ICAST™ System

FIGURE 28

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 August 2001 (16.08.2001)

PCT

(10) International Publication Number
WO 01/058923 A3

(51) International Patent Classification[7]: C12N 15/62, C12P 21/04, G01N 33/53

(21) International Application Number: PCT/US01/00684

(22) International Filing Date: 7 February 2001 (07.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/180,669	7 February 2000 (07.02.2000)	US
09/654,499	1 September 2000 (01.09.2000)	US
09/759,152	16 January 2001 (16.01.2001)	US

(71) Applicant: **TROPIX, INC.** [US/US]; 47 Wiggins Avenue, Bedford, MA 01730 (US).

(72) Inventors: **PALMER, Michelle, A., J.**; 87 Medford Street, Arlington, MA 02174 (US). **GEE, Melissa**; 17 Crescent Avenue, Bedford, MA 01730 (US). **TILLOTSON, Bonnie**; 4 Ripley Road, Belmont, MA 02178 (US). **CHANG, Xiao-jia**; 25 Round Hill Road, Lincoln, MA 01773 (US).

(74) Agents: **KELBER, Steven, B.** et al.; Piper Marbury Rudnick & Wolfe LLP, 1200 Nineteenth Street, N.W., Washington, DC 20036-2412 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report: 24 December 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/058923 A3

(54) Title: IMPROVED SYSTEMS FOR SENSITIVE DETECTION OF G-PROTEIN COUPLED RECEPTOR AND ORPHAN RECEPTOR FUNCTION USING REPORTER ENZYME MUTANT COMPLEMENTATION

(57) Abstract: Methods of detecting G-protein coupled receptor (GPCR) signal activation by employing a GPCR fusion protein comprising an inactive form of an enzyme and an interacting protein partner fusion protein comprising a complementary inactive form of the enzyme contained in the GPCR fusion protein. When the GPCR fusion protein and the protein partner fusion protein bind to one another the two forms of the inactive enzyme associate and produce an active heterodimeric enzyme which is active, permitting detection of the interaction of the GPCR fusion protein with the interacting protein partner fusion protein by measuring the enzyme activity.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/00684

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12N 15/62; C12P 21/04; G01N 33/53
US CL : 435/7.1, 7.2, 69.7, 252.3, 320.1; 536/23.4
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/7.1, 7.2, 69.7, 252.3, 320.1; 536/23.4

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WEST, STN/MEDLINE
search terms: fusion, chimeric?, receptor#, arrestin#, galactosidase, report?, protein-protein interaction#.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	BARAK et al. A β-Arrestin/Green Fluorescent Protein Biosensor for Detecting G Protein-coupled Receptor Activation. The Journal of Biological Chemistry. 31 October 1997, Vol. 272, No. 44, pages 27497-27500, see entire document,	1-25
Y	GUREVICH et al. Arrestin Interactions with G Protein-coupled Receptors. The Journal of Biological Chemistry. 13 January 1995, Vol. 270, No. 2, pages 720-731, especially page 724.	1-25

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search 13 APRIL 2001	Date of mailing of the international search report 23 MAY 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Jayza Bridgers for JOHN D. ULM Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)★

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/00684

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ROSSI et al. Monitoring protein-protein interactions in intact eukaryotic cells by β-galactosidase complementation. Proceedings of the National Academy of Science, USA. August 1997, Vol. 94, pages 8405-8410, see especially Figure 1 on page 8407.	1-25